

# Chromosome Rearrangements and the Evolution of Genome Structuring and Adaptability

Anton Crombach and Paulien Hogeweg

Theoretical Biology and Bioinformatics Group, Utrecht University, Padualaan, Utrecht, The Netherlands

Eukaryotes appear to evolve by micro and macro rearrangements. This is observed not only for long-term evolutionary adaptation, but also in short-term experimental evolution of yeast, *Saccharomyces cerevisiae*. Moreover, based on these and other experiments it has been postulated that repeat elements, retroposons for example, mediate such events.

We study an evolutionary model in which genomes with retroposons and a breaking/repair mechanism are subjected to a changing environment. We show that retroposon-mediated rearrangements can be a beneficial mutational operator for short-term adaptations to a new environment. But simply having the ability of rearranging chromosomes does not imply an advantage over genomes in which only single-gene insertions and deletions occur. Instead, a structuring of the genome is needed: genes that need to be amplified (or deleted) in a new environment have to cluster. We show that genomes hosting retroposons, starting with a random order of genes, will in the long run become organized, which enables (fast) rearrangement-based adaptations to the environment.

In other words, our model provides a “proof of principle” that genomes can structure themselves in order to increase the beneficial effect of chromosome rearrangements.

## Introduction

The sequencing of several eukaryotes and the research that followed in its slipstream lead to important insights. Transposable elements are found to be a source of genetic innovation and to have regulatory functions in many organisms (Biemont and Vieira 2006), gene order is not random (Hurst et al. 2004), and genomes evolve by micro and macro rearrangements (Seoighe et al. 2000; Fischer et al. 2001; Britten et al. 2003). Micro rearrangements include inversions of a couple of genes, single-gene insertions and deletions (indels). Comparative sequence analysis shows that macro rearrangements are localized at the telomeres and centromeres (Eichler and Sankoff 2003; Murphy et al. 2005). It appears that sites on the genome are being reused in the movement and copying of large segments.

In short-term evolution these processes appear to play a role as well. A striking example is given by Dunham et al. (2002). Several repeated experiments were performed where baker's yeast, *Saccharomyces cerevisiae*, was placed in a glucose-limited environment for about 300 generations. Looking at the genomic changes of the resulting strains, they made several observations. Firstly, large chromosome segments are copied and deleted in the majority of strains. Such events are called gross chromosomal rearrangements (GCRs). Secondly, many strains have such huge mutations at the same location on the chromosomes. Dunham et al. (2002) suggest the locations to be fragile sites. Thirdly, although GCRs are abundant, there are also non-GCR ways of adapting to a low-glucose environment (Brown et al. 1998). Adaptation via GCRs has also been observed by Hughes et al. (2000), Infante et al. (2003), and Schacherer et al. (2005), who evolved various deletion mutants. Regaining the function was in half of the cases accompanied by GCRs.

Dunham et al. (2002) found repetitive DNA originating from retroposons at the flanking regions of GCRs. In addition, we know that the transcription of the yeast Ty1

family of retroposons is activated under various stress conditions (Lesage and Todeschini 2005). Both suggest that retroposons are a means of evolutionary adaptation to environmental changes. In a broader perspective retroposons have been implicated in chromosome evolution of yeast (Fischer et al. 2000; Umezu et al. 2002; Hughes and Friedman 2004). In other words, retroposons are linked to genome evolution on different timescales.

Though details need to be filled in, the mechanisms behind GCRs are established. It is known that replication forks tend to stall in regions of repeat elements (e.g., tRNA genes, retroposons, and telomeres) and subsequently cause a double-strand break (DSB) (Cha and Kleckner 2002). From fission yeast, we know that this promotes ectopic recombination (Lambert et al. 2005). Alternatively, Koszul et al. (2004) explains segmental duplications by hypothesizing a role for break-induced replication after DSBs. These modes of homologous recombination result in translocations, deletions, and sister chromatid exchanges (Mieczkowski et al. 2006).

From the above discussed results on short-term adaptation of yeast by several yet reoccurring GCRs, it is tempting to hypothesize that the genome is structured to increase the probability of favorable mutations in alternate environments and hence have an increased adaptability. Such a hypothesis is almost impossible to substantiate experimentally as one cannot rule out that there exists only a tiny set of beneficial mutations and as a consequence selection produces the observed pattern. We therefore apply a computational approach to investigate whether well-established mutational mechanisms could lead to such an outcome.

We define a simple evolutionary model with random mutations in the form of single-gene indels, retroposons that are the source of repeat elements, and DSBs on retroposons that possibly lead to chromosome rearrangements. Given this set of mutational events, we study the evolutionary dynamics in a changing environment. We assume different environments require more or less of certain gene products. For simplicity, we ignore in the present model gene regulation and instead assume adaptation to protein requirements is only through the gene copy number.

In this paradigm system, we show that long-term evolution leads to a structuring of the genome, which in turn leads to faster short-term adaptation to the environment.

Key words: evolution, evolutionary adaptability, genome structure, retroposons, individual-oriented model.

E-mail: a.b.m.crombach@uu.nl.

*Mol. Biol. Evol.* 24(5):1130–1139. 2007

doi:10.1093/molbev/msm033

Advance Access publication February 24, 2007

© 2007 The Authors

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

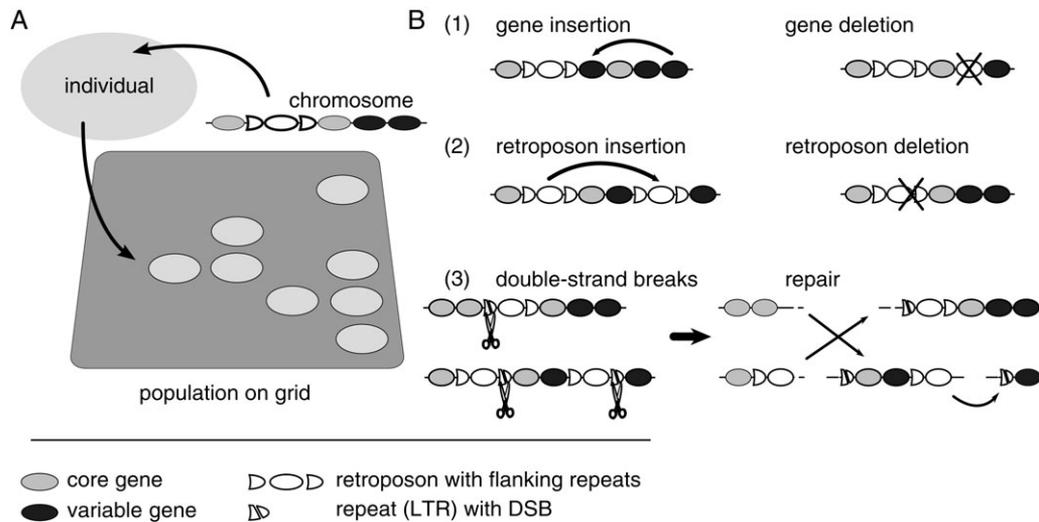


FIG. 1.—Individual-oriented model of retroposon dynamics. (A) The model structure. (B) Three types of mutations: 1) single-gene indels; 2) retroposon copying and removal, removing single LTRs is not shown; and 3) DSBs followed by repair, with rearrangements possibly occurring.

The results suggest that retroposons despite their deleterious effects may have a beneficial effect in the long term.

## Methods

### Model Structure

The model (fig. 1A) consists of an asexually reproducing population of individuals on a grid (spatial structure) adapting to an environment that is homogeneous in space and changing in time. The grid is updated as a standard cellular automaton, that is, all grid cells are updated synchronously. The environment provides the evolutionary goal of the individuals. In its simplest form it switches between 2 target genotypes according to a Poisson process (unless mentioned otherwise  $\lambda = 1.5 \times 10^{-4}$ ). The general idea is that the environment defines the copy number of a subset of genes in the target genotypes.

### Individual

An individual performs 2 actions: reproduce and die. Death happens with a specified probability (fixed at 0.1). Reproduction  $r_i$  requires an empty grid cell to place the offspring. Given such an empty location, the 8 neighboring individuals, called  $nbh$ , compete on basis of their fitness score  $f_i$

$$r_i = \frac{f_i^p}{\max(\sum_{j \in nbh} f_j^p, \theta)}$$

The threshold  $\theta$  (fixed at  $1 \times 10^{-4}$ ) ensures that nothing may happen if there are very few individuals in  $nbh$  or all individuals are very unfit. Given the relative fitness  $r_i$  of each individual in the neighborhood  $nbh$ , one is selected according to the fitness proportional selection scheme. Reproduction itself encompasses copying the genome, mutating, and dividing into 2 daughters. One of the 2 replaces the parent, and the other is placed in the empty grid cell. For simplicity, we only consider asexual reproduction.

In most runs, selection pressure is increased by raising  $f_i$  to a power  $p$  (fixed at 10). It increases the chance that a beneficial mutant spreads in the population, which in turn allows for faster simulations while the results remain qualitatively equivalent. In our results, we discuss the effects of selection pressure in more detail.

### Fitness

An individual holds a genome, which is a single linear “beads-on-a-string” chromosome. In the majority of our simulations, it is initialized with 2 types of genes, 20 “core” and 20 “variable” ones, and 10 retroposons with long terminal repeats (LTRs).

The fitness of an individual  $f_i$  is determined by the environment. The environment switches between 2 states, each associated with an optimal genotype. In one state 1 copy of each of the variable genes is optimal, and in the other 2 copies is optimal. In contrast, 1 copy of each of the core genes is required in both environmental states. Missing any of the genes is lethal, whereas having extra copies results in lower fitness. A biological interpretation is that core genes correspond to genes responsible for essential functions (e.g., cell cycle), and variable genes relate to the ones that process resources (metabolites) from the surroundings.

Fitness is a value in the interval  $[0, 1]$  and defined in terms of a raw score  $s_i$ . Maximizing the fitness amounts to minimizing the raw score.

$$f_i = \begin{cases} 1 - s_i/30, & \text{if } s_i \leq 30, \\ 0, & \text{if } s_i > 30. \end{cases}$$

The raw score is quantified as follows:

$$s_i = D_i + \max(0, t_i - 25).$$

$D_i$  is the distance between the copy number of each gene (both core and variable) and the current optimal genotype. As the retroposon dynamics lack any control mechanism, a penalty is added to the gene distance if  $i$  has more

than 25 retroposons ( $t_i$  is the number of retroposons in the genome). A penalty on the number of single LTRs is included in a genome size penalty (threshold size 250). However, it is generally not applied during runs and therefore left out of the formulas.

We also perform simulations with an extra group of 20 variable genes that follow an additional, independent environmental cue. This creates a setting with 4 different environmental states that the individuals adjust their 2 subsets of variable genes to.

### Mutational Events

At reproduction, the chromosome is duplicated, after which 3 types of mutational events may occur on the diploid genome (fig. 1B). The first type is “gene indels”: gene insertion and deletion. The former is the act of copying a gene and placing it at a random position in the genome, though it is never inserted in between a retroposon and its LTRs. The latter is deleting a gene. The second type is “retroposon dynamics”: retroposon insertion and deletion and LTR deletion. Insertion is copying a retroposon (including the flanking LTRs) and inserting it at a random position in the genome. Deletion is always done via single-strand annealing, which leaves a single LTR. Such a single element (i.e., one that is not next to a retroposon) can be removed as well. The third type of mutation is GCRs, which happen through DSBs at LTRs. These DSBs are repaired by randomly reattaching chromosome segments to each other, with the constraint that the beginning and end of each chromosome are kept as such. In other words, the first and last segment of a DSB-damaged chromosome are the first and last of a recombined chromosome. At least 1 DSB per chromosome is needed for a rearrangement to occur (swapping tails). During this process no chromosome segments are lost; however, the resulting 2 chromosomes may be of unequal length and/or content. Thus, as each daughter cell receives a chromosome, they may have deletions and/or translocations of chromosome segments.

### Ancestor Tracing

During a simulation each individual has its own unique identification and knows its parent. This enables us to reconstruct genealogies: one of the best individuals at the end of the simulation is selected and all its ancestors are traced back to the start. By recording genomes and which mutations occur along this lineage, we are able to look at the mutational mechanics in detail.

## Results

### Two Typical Runs

In the evolutionary runs, the starting point is that retroposons have successfully invaded the population. We discuss 2 typical runs: a run of  $1 \times 10^6$  time steps with 1 set of 20 variable genes per individual and one of  $2 \times 10^6$  time steps with 2 sets of 20 variable genes. In both runs a homogeneous population is initialized on a  $100 \times 100$  grid and subjected to a changing environment.

### One Group of Variable Genes

The run is shown in figure 2A. Each time the environment switches, the average gene distance jumps to 20 (maximal distance) and the population adapts to the new environment.

In figure 2B, the close-up shows a small GCR with a net distance gain of 4, followed by single-gene indels. The variable genes make a little jump due to the GCR and then slowly increase to a double copy number (shaded area). We see a genotype with a few extra core genes spread through the population immediately after the environmental switch. This hitchhiking of core genes is caused by the small GCR. It is interesting to note that GCRs are readily applied, even though genes are randomly ordered on the genome and the extra core copies need to be removed again. The close-up at the end of the run (fig. 2C) shows a rather different style of adapting. Mutations by GCRs spanning a full set of variable genes cause the population to adapt extremely fast. The doubling of the variable genes still shows a few hitchhiking core genes, which are subsequently deleted.

These observations indicate a long-term process of genome restructuring. To quantify the ordering of genes on the chromosomes, we look at all adjacent pairs of genes, while ignoring retroposons and LTRs. We take Cramer’s phi coefficient (Cramer’s phi is derived from the chi-square statistic, in formula:  $\phi = \sqrt{\chi^2 / (N^*(k - 1))}$  with, in our case,  $N$  the total number of genes and  $k$  the number of variable groups) of these pairs, which gives a value in the range [0,1]. The higher the phi coefficient, the higher the degree of clustering of genes by type (core or variable).

In figure 3A, the average organization of genomes in the population is shown. The first third of the run the genomes do not show any clustering, after which the population evolves toward a high degree of organization ( $\phi = 0.76$ ). This level is kept, although it sometimes drops slightly (e.g., at  $t = 6500$  and  $t = 9800$ ). The temporary declines are explained by the fact that there are only very few sequences in gene-order space with a high phi coefficient, and apparently, the indirect selection is not strong enough to maintain them. The main cause is retroposons copying themselves through the genome, creating alternative break points, and hence, via rearrangements the gene order is randomized to a small degree.

The clustering of genes, separated by strings of LTRs is clearly visible if we look at the genomes in figure 4. Figure 4B shows from left to right the core genes and 2 clusters of variable genes. Apparently, the individuals are increasing the probability of DSBs at certain locations in their genome and therewith enable fast adaptations to the environment. For instance, a scenario of a rearrangement would involve a DSB in the middle of one chromosome and at the right-hand side of the other chromosome. In this manner swapping the right-hand tails would result in copying half of the variable genes with only 1 or 2 core genes hitchhiking.

### Two Groups of Variable Genes

The evolution to an organized genome in which only 2 types of genes are present, seems a rather simple task. We

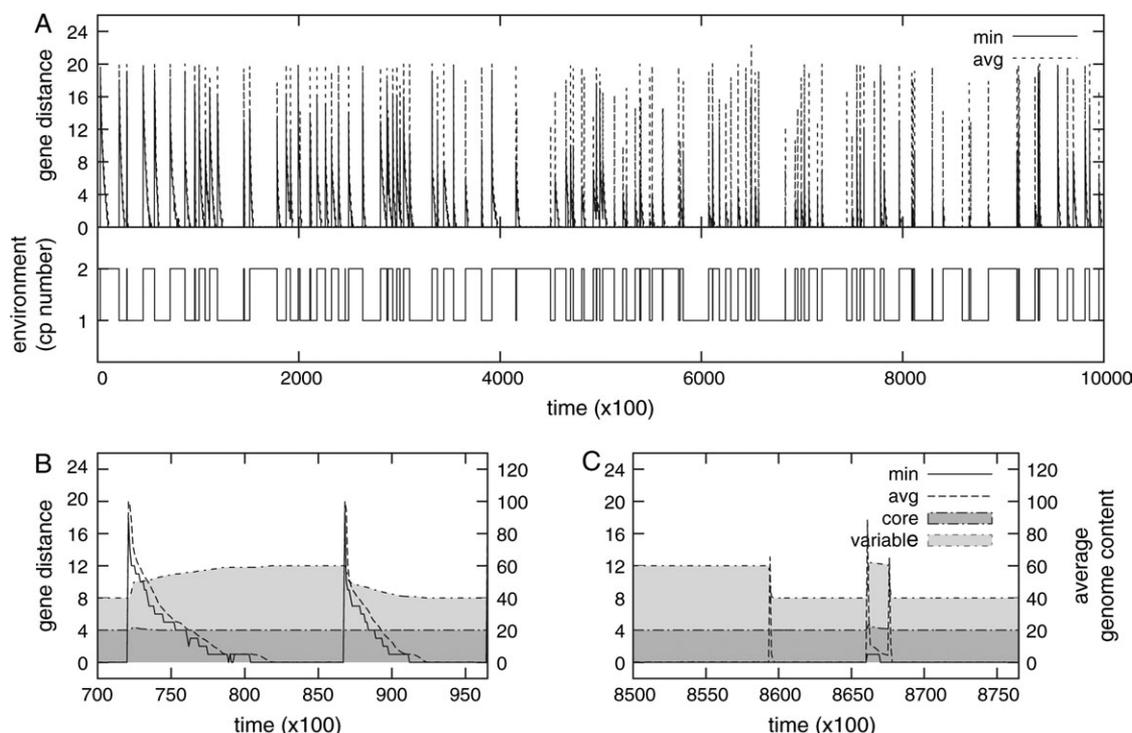


FIG. 2.—Typical run. Parameters (per gene, retroposon, and LTR): single-gene copy and removal =  $0.5 \times 10^{-5}$ , retroposon copy, removal, and LTR removal =  $1 \times 10^{-5}$ , and DSB repair =  $6 \times 10^{-4}$ . Note that at least 2 breaks are needed for a chromosome rearrangement. The environment switches at  $\lambda = 1.5 \times 10^{-4}$ . (A) The top part shows the average (avg) and minimum (min) gene distance of the population. Every environment change, shown at the bottom, is accompanied by a peak in the gene distance. (B) and (C) are close-up graphs of (A). Both show on the left y axis the average and minimum gene distance and on the right axis (shaded) average genome content (i.e., the number of core and variable genes). (B) Two hundred time steps at the start showing a gradual decrease of gene distance by small steps (indels). (C) Two hundred time steps at the end with large and immediate decreases of gene distance. The fast mutation is also observed in the shaded area of variable genes, whereas the area of core genes remains constant.

show that this behavior can be extended to multiple groups of variable genes. This creates a more complicated task where in one environment more of one group of genes is needed, in another more of the other group of variable genes and in yet other environments both or none are needed in more than one copy. For  $2 \times 10^6$  time steps, we have run a population with individuals having 20 core genes and 2 groups of 20 variable genes. The environment has an independent signal for each variable gene group, that is, a total

of 4 environmental states, both with a probability of change  $\lambda = 1 \times 10^{-4}$ .

We observe qualitatively similar behavior in the adaptations to a new environment as the “one-group” case (data not shown). For instance, core genes hitchhike with both variable groups and the variable groups show hitchhiking among each other as well. Again, indirect selection causes genes to cluster by type. In figure 3B Cramer’s phi is plotted as the measure of average organization in the population. If

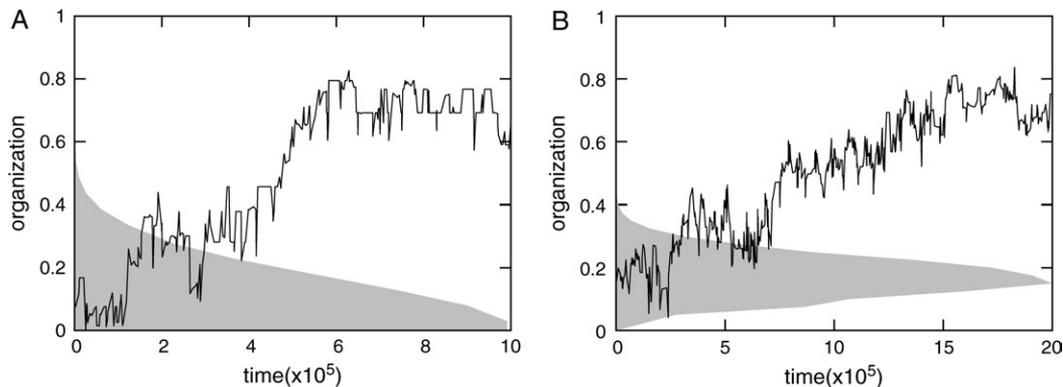


FIG. 3.—Average genome organization in the population. (A) One set of variable genes. The run of figure 2 is shown. (B) Two sets of variable genes. In both figures simulation time is set against average organization of the population. The latter is expressed as Cramer’s phi coefficient. To demonstrate the significance, we show shaded areas that give, in the horizontal direction, the frequency distribution of organization from a sample of one million randomly generated genomes. Note that the second simulation is twice as long.

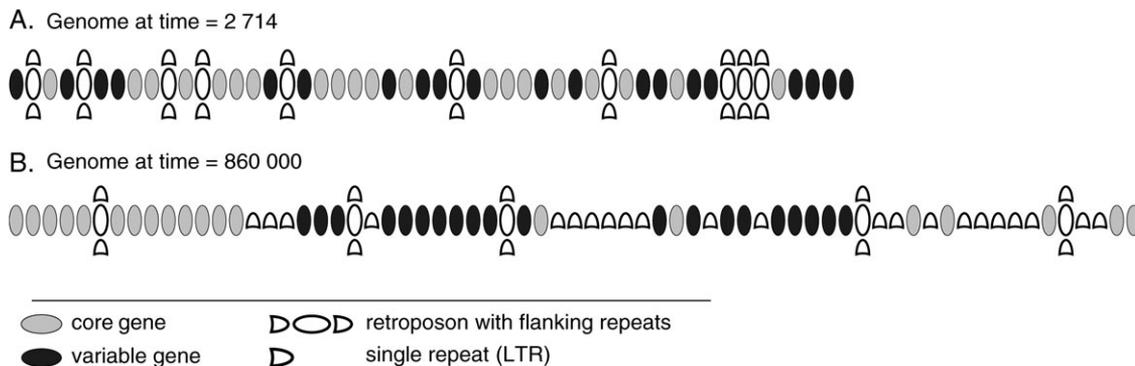


FIG. 4.—Gene order in 2 randomly picked individuals, early and late in the simulation. The genomes are taken from the run of figure 2. Genome A shows no organization, in contrast to genome B, which displays a clear clustering of the 2 types of genes and an increase in single LTRs.

we compare the curve with the one-group case, it is remarkable that similar levels of organization are reached in this more complex case and that it takes no more than twice the amount of time.

The majority of the results that we discuss next is based on the behavior of the one-group case as the simulations are rather computationally intensive. The “two-group” case is considered to be a strong indication of the generality of these results within our framework.

Mutational Dynamics

We study the one-group run in more detail. In the top graph of figure 5 the running average of single-gene indels in the entire population is shown. Deleterious mutations make up the bulk and have a fairly constant rate throughout the run, whereas the number of beneficial indels decreases. The rearrangements show a different behavior. If we look at the ancestor trace, it provides us with a detailed view on the dynamics of the “beneficial” mutational mechanics. The bottom graph (fig. 5) shows the usage of GCRs in the ancestor trace. In contrast to the change in frequency such as the indels show, the rearrangements keep a rather constant rate. However, GCRs become more effective as the simulation progresses.

The above observations are reinforced by table 1. If we take the first  $3 \times 10^5$  time steps (table 1, first row), we know genomes are not yet organized. Hence, we observe many single-gene indels in the ancestors and they are mostly beneficial ones. Obviously few mutations are deleterious due to the strong selection. Even though we do not have organization, GCRs are applied, as the number of beneficial ones more or less equals the overall occurrence (table 1, third row). This is in accordance with the previous observation that rearrangements occur at about the same rate during the run. The difference with the population average shows that virtually all beneficial GCRs occur in the ancestor lineage.

After  $6 \times 10^5$  time steps, the population consists of organized individuals (table 1, second row). Compared with the beginning of the run, single-gene insertions have nearly disappeared. Gene removals have decreased drastically too, but still occur. The explanation is that, as mentioned in the previous section, core genes tend to hitchhike on rearrangements and subsequently 1 of the 2 genes is removed via a single-gene deletion. GCRs are applied less often than

on average in the simulation, but this is compensated by having GCRs that span more variable genes. The slight rise of deleterious GCRs seems paradoxical, but is caused by mutations that by accident anticipate an environmental switch. These mutations are categorized as deleterious, but change to beneficial within the individual’s lifetime. Such rearrangements of anticipatory nature strengthen the idea that there are locations on the genome with a higher probability of a DSB. Correcting for these anticipatory GCRs, the rate of deleterious ones in the ancestor lineage is  $0.49 \times 10^{-4}$  per reproduction, which is lower than the population average.

We conclude that the population is forced to deal with retroposons and the frequent rearrangements they cause. One would expect to see the retroposons removed by selection because of the higher death rate they cause. Indeed, chromosomal rearrangements are almost always bad as seen from the 3 orders of magnitude difference between advantageous and deleterious GCRs in the population (table 1,

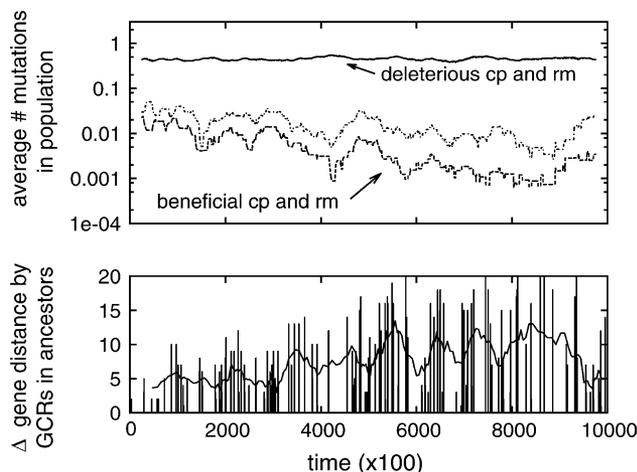


FIG. 5.—Mutations in the run of figure 2. In the top graph the population running average (window size 100001) of beneficial and deleterious single-gene indels, that is, copy (cp) and remove (rm), is shown. The bottom figure shows fitness gains expressed as  $\Delta$  gene distance made by GCRs in the ancestor trace: the decrease in distance to the target genotype. The vertical lines show the actual gains made by each GCR event, whereas the running-average curve (window size 9) shows the trend of increasing gains.

**Table 1**  
**Mutations in the Ancestor Lineage and Population**

Interval ( $\times 100$ )	Number of Individuals	Mutations ( $\times 10^{-4}$ ) Ben   Del					
		Copy		Remove		GCR	
0–3,000	64,386	16.93	0.15	34.17	0.77	10.56	0.62
6,000–10,000	82,466	0.12	0	9.34	0	9.82	1.21 (0.49)
0–10,000	209,720	6.77	0.05	18.60	0.29	10.11	0.95
0–10,000	Pop $\approx 9 \times 10^9$	0.0064	0.50	0.0184	0.49	0.0026	0.71

NOTE.—Average beneficial (Ben) and deleterious (Del) mutations per individual with the individuals being from the ancestor trace or the population (4th row). In the 1st interval individuals have not organized their genomes yet, in the 2nd interval they have. The 3rd interval is the entire simulation (see fig. 2). In the 2nd row the deleterious gross chromosomal rearrangements (GCR) value in parentheses is corrected for anticipatory rearrangements. As a reference to the ancestor trace, the bottom row shows average mutations for the population (Pop) in an entire run.

bottom row). But by reordering their genomes, individuals utilize the effects of GCRs and gain a novel (fast) way of adapting, which overrules the negative effects of retroposons and their flanking LTRs. The mechanism of restructuring the genome depends on GCRs, the selective loss of surplus genes, and the reordering due to single-gene indels.

### Invasion of Retroposons

Invasion and maintenance of retroposons in a sexually reproducing population has been explained in terms of transposition and recombination (Rouzic and Capy 2005). It is argued that in a clonal population selection is necessary to explain the presence of retroposons (Edwards and Brookfield 2003). In our model, we have seen maintenance of retroposons through the indirect selection for evolvability. However, we expect such selection to be minimal during an invasion, as the genomes are still unorganized.

We study the paradox of fixing retroposons in a host genome by means of invasion simulations. A GCR-enabled population is introduced in an “optimal” indel-only population for different initial population sizes and retroposon numbers. The settings of a typical run are taken with one alteration. The indel-only individual mutates fast if it needs to adapt, yet it hardly mutates if it is perfectly fit. Such behavior is accomplished by evolving the mutation rates as well: during reproduction an individual may mutate (probability  $5 \times 10^{-4}$ ) its single-gene mutation rates. The new value is drawn from a uniform distribution in the range  $[0, 2.5 \times 10^{-4}]$ . The upper bound is above the maximal mutation rate ( $2.0 \times 10^{-4}$ ) for which adaptation to the environment can be maintained. Thus, we assume a worst-case scenario to study the invasion dynamics.

If we let a fraction of the population of 0.1 ( $\approx 1000$  individuals) have 5 retroposons per genome, 32/400 runs result in a successful takeover. Increasing the number of individuals to 2000 gives success in 52/400. Naturally, the more the GCR-enabled individuals, the higher the probability of taking over the population, though even a majority does not guarantee a successful invasion (data not shown). The interesting observation is that for smaller numbers of invading individuals, we still have successful invasions. If we reduce the fraction of individuals with retroposons to 0.05 ( $\approx 500$  individuals), 8/400 invasions succeed. Taking a smaller patch of 100 individuals gives success in 6/400.

We conclude that although retroposons do not give an inherent selective advantage in an unordered genome, on an evolutionary timescale it seems fairly easy for them to in-

vade a population based on the positive effect they have on evolvability if they happened to be at the appropriate location in the genome. Thus, we have a convenient mechanism to introduce repeat elements into a genome.

### Rate of Environmental Change

An important parameter in our model is the rate of switching from one state to another. In our typical runs, we used  $\lambda = 1.5 \times 10^{-4}$  for the one variable–group simulation and  $\lambda = 1.0 \times 10^{-4}$  (per environmental signal, thus effectively  $\lambda = 2.0 \times 10^{-4}$ ) for the more challenging two-group case. As we see clearly in figure 6A, in these ranges of  $\lambda$  it is usually possible for the population to fully adapt to the current environment.

When we set  $\lambda = 1 \times 10^{-3}$ , the global behavior changes. In the first half (fig. 6B), often the population does not reach the optimal genotype. Instead the fast fluctuations of the environment cause the population to average over the 2 environmental states. However, there is still a bias toward low distances. The bias and as a consequence, the small adaptations to each environmental switch are sufficient to trigger the rise of gene ordering by group. In the second half of the run, the population consists of organized genomes. Even more interesting is that on a population level, we always have 2 genotypes, that is, we have population-based diversity where the opposite genotype, which is present at almost 5%, is constantly being generated from the current optimal genotype.

We increased  $\lambda$  to  $3 \times 10^{-3}$  (fig. 6C). The population basically experiences an average environment and settles at an average gene distance of 10, half of a variable-group size. Yet, if we start the run with organized individuals, the gene ordering is kept and the population switches from one to the other environment, behaving like figure 6B. There also appears to be a trend to increase the average number of LTRs to cope with such a fast switching rate (data not shown).

The second half of these 3 runs is summarized in figure 6D. For populations with organized genomes, the peaks at distance 0 and 20 are clearly visible. The latter is more prominent for the run from figure 6B, which emphasizes the idea of population-based diversity. The “averaging” run (fig. 6C) shows a characteristic curve with a mean distance around 10.

### The Role of LTRs

The interesting behavior we see in figure 6B is that despite the near absence of adaptation in the first half,

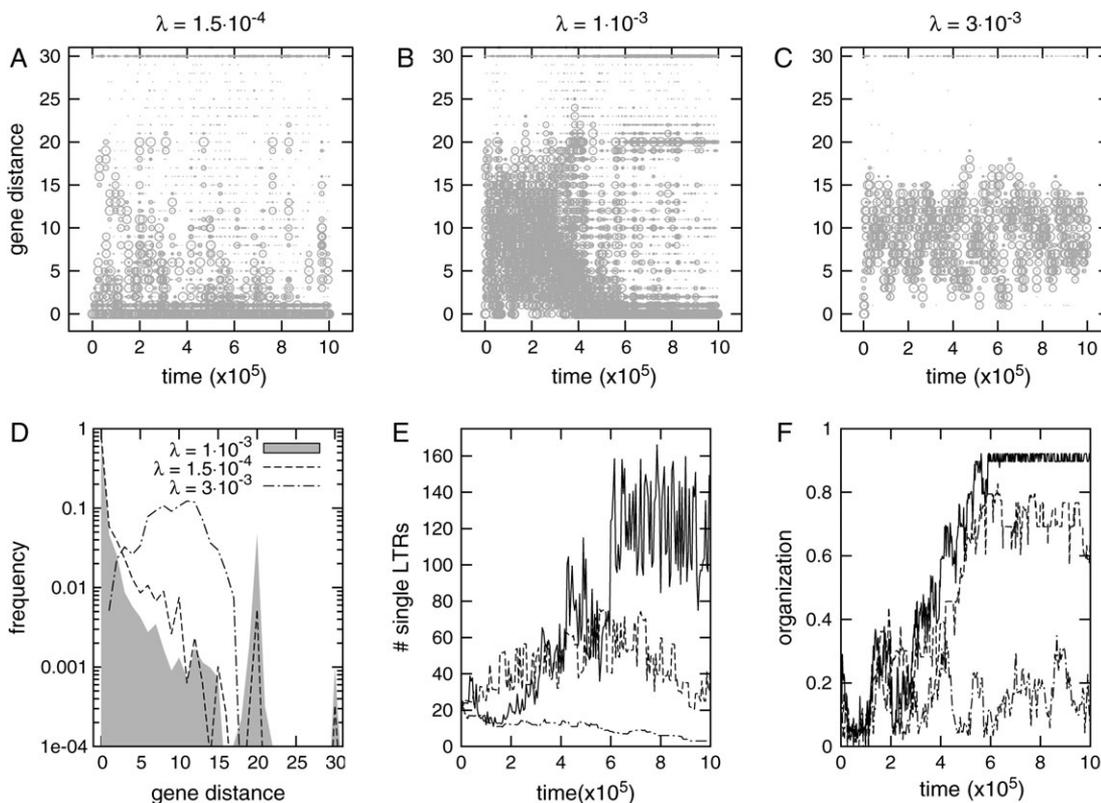


FIG. 6.—Behavior for different rates of environmental change ( $\lambda$ ). (A), (B), and (C) show the distribution of distances to the optimum (0) during the run. Populations are sampled at regular intervals (period of 10,000 in A, 1,000 in B and 4,000 in C) and also at environmental switching in (A). The size of each circle represents (logarithmically) how many individuals have such a distance to the optimal genotype. (D) The frequency distribution of the gene distances during the second half of the runs. The intermediate rate of change is shaded. (E) The number of single LTRs for each of the 3 runs. Please note that the upper curve depicts the intermediate value of environmental change, whereas the bottom curve is the highest rate of change. (F) The average organization in the populations expressed as Cramer's phi. Upper, middle, and lower curve have the same  $\lambda$  values as in (E).

organization, and therewith adaptation, may still arise and even reach higher levels than in our 2 typical runs, as discussed above. In most runs this is accomplished by both retroposon dynamics and LTRs. Figure 6B is a special case as due to a stochastic fluke retroposons are purged from the population before any truly organized genomes dominate the dynamics. Only single LTRs are left and as the only mutation LTRs can have is deletion from the genome, we would expect them to disappear. However, due to positive indirect selection of small beneficial GCRs, the removal of LTRs is extremely slow. Such GCRs occur already in nonorganized genomes and are often fixed in the population (i.e., the mutant takes over the population).

As organization starts to develop, the number of single LTRs per individual increases (see fig. 6E). The growth is accomplished by rearrangements. LTRs tend to cluster between core and variable genes, thus increasing the probability of creating a fit daughter for the opposite environment. GCRs resembling unequal crossover then may enlarge or shorten these sections of single LTRs. Eventually the organization stabilizes (fig. 6F), whereas the number of single LTRs fluctuates rapidly around a mean of approximately 120.

Apparently the reordering is the result of gene indels and the maintenance of single LTRs by indirect selection. This leads to the conclusion that retroposons are not nec-

essary for genome structuring, only repetitive elements like the LTRs are. This is in accordance with simulations started with only single LTRs (data not shown).

#### The Role of Selection

The actual value of  $\lambda$  for which we observe these modes of behavior depends on the selection power  $p$  (see Methods). We relax selection pressure if we choose low values of  $p$  ( $p = 2$  for example). Consequently there is only a small beneficial effect of having one distance less, resulting in a slow adaptation of the population to the optimal genotype. Thus, to observe the evolution of adaptability (reordered genomes), a low  $\lambda$  is needed and due to a tighter error threshold, it is necessary to lower mutation rates as well. On the other hand, as  $p$  increases the error threshold is relaxed, hence higher values of  $\lambda$  can be chosen. This enables us to speed up simulations. Except for the timescale there is no qualitative difference in the obtained results.

#### Parameter Sensitivity

The organization as described in the typical runs is a robust phenomena. However, for identical parameter values the onset of organization ranges from soon after the start to halfway through the run.

The robustness is further explored by starting at different mutation rates, by removing any spatial patterns that may influence the results and by introducing extra penalties per retroposon. Besides initializing at various rates, we let the individuals evolve their mutation rates. The rates may change during a simulation, and each changes with small steps. We have 3 different step sizes, 1 for each type of mutation rate (indel step  $1.5 \times 10^{-7}$ , retroposon  $1 \times 10^{-7}$ , and DSB repair  $1 \times 10^{-3}$ ). By allowing mutating mutation rates, we broaden our view on parameter space with a limited set of runs. For instance, if we start with the rates of a typical run and these rates would evolve to very different values and behavior, we could say that the behavior of a typical run is an artifact of “forced” rates, not natural behavior.

### *Changing Mutation Rates*

We vary each type of mutation events by decreasing or increasing the rate, whereas the rest of the parameters are initialized as for the typical run (see fig. 2). For each setting, we have run 3 simulations, which we discuss shortly. For 3 times lower DSB rates ( $2 \times 10^{-4}$ ), organization develops more slowly simply because less breaks occur. An almost twice as high rate of DSB repair ( $10 \times 10^{-4}$ ) does not show any qualitative difference. Ten times lower retroposon rates ( $0.1 \times 10^{-5}$ ) make the mobile genetic elements more vulnerable to stochastic fluctuations. In 2 of the 3 runs, retroposons are removed from the population, but single LTRs are kept. Individuals manage to gain and keep organization in these runs, albeit less stable. For single-gene rates, we observe that more than twice higher rates ( $1.2 \times 10^{-5}$ ) result in better organization, even above 0.80. Because the organization scale is not linear, but represents half a normal distribution, such levels of organization are extremely rare. They are due to the increased reordering effect of single-gene mutations. Indeed, a 5 times lower indel rate ( $0.1 \times 10^{-5}$ ) than in a typical run results in a lower level of organization, around 0.6, within an equal amount of simulation time.

### *Retroposon Penalty*

Throughout a run, retroposons cause many lethal mutants by generating erroneous chromosome rearrangements. In other words, retroposons increase the death rate of individuals. We examine if the addition of a fitness penalty per retroposon (0.04 extra gene distance per retroposon) instead of a penalty if a threshold of retroposon copies is exceeded, is a viable alternative. In this scenario, retroposons are rather quickly removed from the population. Yet, individuals still reshuffle their genomes. They apply the trick of using single LTRs, as described above.

### *Structural Stability*

In our typical run, variable genes comprise a large part of the genome, either 20 or 40 genes, compared with 20 core genes. The question arises if the ratio core-variable genes influences the evolution of organization. We have performed simulations with more core than variable genes, such that the number of core genes (50 genes) is always more than the variable ones. The runs show that genomes still develop organization.

Another test of the structural stability of our model is the introduction of a special element: the centromere. We add the constraint that each chromosome has to have 1 centromere, whereas both having more than one (dicentric) and missing the centromere (acentric) is lethal. The simulations resulted in organized genomes, with the centromere being pushed in the direction of the beginning or end of the chromosome.

## **Discussion**

The concept that “some mutations are more equal than others” is rather controversial. Generally, mutations are considered random events, with selection acting on them. However, our results show that some classes of mutations may occur preferentially, and furthermore, these mutations are likely to be advantageous. In our paradigm model, we take random, well-established mutational events and show in our results that some (phenotypic) mutations are being favored. We observe that genomes restructure their gene order such that chromosome rearrangements are very likely to produce the target genotype of an alternate environment. However, it is not the only possible behavior. Depending on the rate of change of the environment and the previously achieved degree of ordering of the genomes, we may see the evolution of an average genotype with low fitness that “integrates” over the environmental states.

There are 3 assumptions of our model we would like to discuss. Firstly, we make a worst-case assumption by using a very simple genotype–phenotype relation, in which each alteration in gene copy numbers affects an individual’s fitness. In contrast, experimental findings show that the detrimental effects of having an extra gene copy are not so large (Wilke and Adams 1992). Often it is compensated for on the gene regulation or metabolic level.

Secondly, our method for resolving DSBs only approximates the *in vivo* mechanism. We ignore explicit homologous repair as an option for resolving DSBs, and we assume randomness in concatenation of chromosome parts. *In vivo*, DSB repair is not completely arbitrary, but there are many blanks in our knowledge. Therefore, we apply our random repair protocol. For the parameter ranges we investigated, if a GCR occurs, on average 2 breaks occur in the genome. This allows for simple relocation of a chromosome segment; an entire randomization of the chromosomes is not observed. We also assume that retroposons insert themselves at random positions, although Ty families in baker’s yeast have a preference for certain genome locations (Lesage and Todeschini 2005).

Thirdly, if selection is reduced, drifting by gene indels takes place. This is a source of randomizing the gene order. Yet, we mainly use a strong selection pressure, hence mutational drift does not occur. It means that keeping the organization in a constant environment becomes rather straightforward. Therefore, we may not extrapolate into environmental switching at low rates, unless we lessen selection pressure.

Another model feature worth mentioning is that the individuals evolve in an explicit spatial setting with local competition. By performing simulations in which the population is mixed at each time step, we established that

spatial pattern formation does not play a role in our results (data not shown). This is in agreement with previous results that show well-mixed populations employ evolutionary adaptation and mutational priming (Hogeweg 2005) as their main strategy.

Pepper (2003) studied gene linkage given unequal crossing over and inversions. While ignoring retroposons or repeat elements, he finds a clustering of genes too. In his discussion, he proposes an evolutionary scenario with, on a longer timescale, a positive feedback between rearrangements and gene linkage, which resembles our results. The long-term result of adaptability is only indirectly selected for in our simulations as the fitness criterion, the short-term advantage, does not contain any reference to the gene linkage we observe at the end. There is selection on lineage level leading to higher levels of gene ordering and a higher degree of adaptability. What actually happens is that the genomes restructure their mutational landscape with the retroposons. They develop a “coding scheme” that allows them to change swiftly between 2 phenotypes in a genotypic manner, also known as mutational priming. In principle, repetitive elements such as LTRs are sufficient for generating and sustaining the organization. Retroposons may be regarded as a vector for introducing such elements into a genome. At first the mobile elements only hinder their host by causing chromosomal rearrangements. Such arrangements are in almost all cases deleterious. However, retroposons can establish themselves by providing individuals with the opportunity of using any bit of clustering of genes via GCRs to adapt to a new environment. Single-gene indels, together with retroposon indels then amplify the grouping of genes by their type.

In the Introduction, we already stated that eukaryotes have a nonrandom gene order. This organization of the genome is observed in terms of clusters of coexpressed genes (Singer et al. 2005), clustering of genes encoding subunits of protein complexes (Teichmann and Veitia 2004) and functionally related genes (Hurst et al. 2004). All are clustering seen in the light of gene expression. In yeast, bidirectional promoters also give a direct relation for coexpression between gene pairs (Cohen et al. 2000). How the eukaryotic genome ordered itself is an open question. Naturally, the null hypothesis is that it is not under selection. As gene expression is a noisy process, it could just be an effect of expression leaking (Hurst et al. 2004). Or it could be a side effect of mutational dynamics. For instance highly expressed genes are within open chromatin, which in turn facilitates invasion of new genes (Hurst et al. 2004).

With our model, we add a hypothesis to the “gene expression” and “open chromatin” ones, which may be labeled as the “evolvability” hypothesis, that is, gene ordering evolves as a consequence of chromosome rearrangements and increases adaptability. Interestingly, in yeast a lot of remnants of retroposon activity are observed, which we could now hypothesize to still have a functional role in evolution.

In the evolutionary experiments with yeast (Brown et al. 1998; Ferea et al. 1999; Dunham et al. 2002; Schacherer et al. 2004), it is clear that both evolutionary and regulatory adaptation play a role. First, in most experi-

ments GCRs occur, and second, different GCRs and cases without rearrangements appear to lead to similar changes in gene expression. In addition, the majority of genes that are over- or underexpressed are not located on the duplicated or deleted chromosome segments. Thus regulation creates a far more complicated mapping from genome to phenotype than we have considered. We should note that the amount of organization we observe is much larger compared with, for instance, yeast. At present, we cannot rule out that the GCR observed in the evolutionary experiments only cause minor improvements and resemble the ones we frequently observe for randomly ordered genomes. In future work, we aim to investigate the interplay between evolutionary adaptation as studied here and regulatory adaptation by extending our model with gene regulation.

Our model provides a “proof of principle” that genomes can structure themselves so as to utilize the beneficial effects of chromosome rearrangements. In short, we provide a simple but sufficient model that shows evolution of evolutionary adaptability.

### Acknowledgments

We thank Otto Cordero and Nobuto Takeuchi for their suggestions and discussions. This research is funded by the Computational Life Science program of the Netherlands Science Organization (NWO) under grant number 653-100-001.

Funding to pay the Open Access publication charges for this article was provided by NWO grant 653-100-001.

### Literature Cited

- Biemont C, Vieira C. 2006. Genetics: junk DNA as an evolutionary force. *Nature*. 443:521–524.
- Britten RJ, Rowen L, Williams J, Cameron RA. 2003. Majority of divergence between closely related DNA samples is due to indels. *Proc Natl Acad Sci USA*. 100:4661–4665.
- Brown C, Todd K, Rosenzweig R. 1998. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. *Mol Biol Evol*. 15:931–942.
- Cha RS, Kleckner N. 2002. ATR homolog Mec1 promotes fork progression, thus averting breaks in replication slow zones. *Science*. 297:602–606.
- Cohen BA, Mitra RD, Hughes JD, Church GM. 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet*. 26:183–186.
- Dunham M, Badrane H, Ferea T, Adams J, Brown P, Rosenzweig F, Botstein D. 2002. Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA*. 99:16144–16149.
- Edwards RJ, Brookfield JFY. 2003. Transiently beneficial insertions could maintain mobile DNA sequences in variable environments. *Mol Biol Evol*. 20:30–37.
- Eichler EE, Sankoff D. 2003. Structural dynamics of eukaryotic chromosome evolution. *Science*. 301:793–797.
- Ferea T, Botstein D, Brown P, Rosenzweig R. 1999. Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc Natl Acad Sci USA*. 96:9721–9726.
- Fischer G, James SA, Roberts IN, Oliver SG, Louis EJ. 2000. Chromosomal evolution in *Saccharomyces*. *Nature*. 405:451–454.

- Fischer G, Neuvéglise C, Durrens P, Gaillardin C, Dujon B. 2001. Evolution of gene order in the genomes of two related yeast species. *Genome Res.* 11:2009–2019.
- Hogeweg P. 2005. Self-organisation and evolution of social systems. In: Hemelrijk CK, editor. *Interlocking of self-organisation and evolution*. Cambridge: Cambridge University Press. p. 166–189.
- Hughes AL, Friedman R. 2004. Transposable element distribution in the yeast genome reflects a role in repeated genomic rearrangement events on an evolutionary time scale. *Genetica.* 121:181–185.
- Hughes TR, Roberts CJ, Dai H, et al. (12 co-authors). 2000. Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat Genet.* 25:333–337.
- Hurst LD, Pál C, Lercher MJ. 2004. The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet.* 5:299–310.
- Infante JJ, Dombek KM, Rebordinos L, Cantoral JM, Young ET. 2003. Genome-wide amplifications caused by chromosomal rearrangements play a major role in the adaptive evolution of natural yeast. *Genetics.* 165:1745–1759.
- Koszul R, Caburet S, Dujon B, Fischer G. 2004. Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J.* 23:234–243.
- Lambert S, Watson A, Sheedy DM, Martin B, Carr AM. 2005. Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier. *Cell.* 121:689–702.
- Lesage P, Todeschini AL. 2005. Happy together: the life and times of Ty retrotransposons and their hosts. *Cytogenet Genome Res.* 110:70–90.
- Mieczkowski PA, Lemoine FJ, Petes TD. 2006. Recombination between retrotransposons as a source of chromosome rearrangements in the yeast *Saccharomyces cerevisiae*. *DNA Repair (Amst).* 5:1010–1020.
- Murphy WJ, Larkin DM, van der Wind AE, et al. (25 co-authors). 2005. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science.* 309:613–617.
- Pepper JW. 2003. The evolution of evolvability in genetic linkage patterns. *Biosystems.* 69:115–126.
- Rouzic AL, Capy P. 2005. The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. *Genetics.* 169:1033–1043.
- Schacherer J, de Montigny J, Welcker A, Souciet JL, Potier S. 2005. Duplication processes in *Saccharomyces cerevisiae* haploid strains. *Nucleic Acids Res.* 33:6319–6326.
- Schacherer J, Tourrette Y, Souciet JL, Potier S, Montigny JD. 2004. Recovery of a function involving gene duplication by retroposition in *Saccharomyces cerevisiae*. *Genome Res.* 14:1291–1297.
- Seoighe C, Federspiel N, Jones T, et al. (20 co-authors). 2000. Prevalence of small inversions in yeast gene order evolution. *Proc Natl Acad Sci USA.* 97:14433–14437.
- Singer GAC, Lloyd AT, Huminiecki LB, Wolfe KH. 2005. Clusters of co-expressed genes in mammalian genomes are conserved by natural selection. *Mol Biol Evol.* 22:767–775.
- Teichmann SA, Veitia RA. 2004. Genes encoding subunits of stable complexes are clustered on the yeast chromosomes: an interpretation from a dosage balance perspective. *Genetics.* 167:2121–2125.
- Umezu K, Hiraoka M, Mori M, Maki H. 2002. Structural analysis of aberrant chromosomes that occur spontaneously in diploid *Saccharomyces cerevisiae*: retrotransposon Ty1 plays a crucial role in chromosomal rearrangements. *Genetics.* 160:97–110.
- Wilke CM, Adams J. 1992. Fitness effects of Ty transposition in *Saccharomyces cerevisiae*. *Genetics.* 131:31–42.

Diethard Tautz, Associate Editor

Accepted February 16, 2007