Transcriptional Mutagenesis Prevents Ribosomal DNA Deterioration: The Role of Duplications and Deletions

Enrico Sandro Colizzi[1,2]* and Paulien Hogeweg[3]

[1]Origins Center, 9747AG, Groningen, the Netherlands

[2]Mathematical Institute, Leiden University, the Netherlands

[3]Theoretical Biology and Bioinformatics, Utrecht University, The Netherlands

*Corresponding author: E-mail: e.s.colizzi@math.leidenuniv.nl.

Accepted: October 24, 2019

Abstract

Clashes between transcription and replication complexes can cause point mutations and chromosome rearrangements on heavily transcribed genes. In eukaryotic ribosomal RNA genes, the system that prevents transcription–replication conflicts also causes frequent copy number variation. Such fast mutational dynamics do not alter growth rates in yeast and are thus selectively near neutral. It was recently found that yeast regulates these mutations by means of a signaling cascade that depends on the availability of nutrients.

Here, we investigate the long-term evolutionary effect of the mutational dynamics observed in yeast. We developed an in silico model of single-cell organisms whose genomes mutate more frequently when transcriptional load is larger. We show that mutations induced by high transcriptional load are beneficial when biased toward gene duplications and deletions: they decrease mutational load even though they increase the overall mutation rates. In contrast, genome stability is compromised when mutations are not biased toward gene duplications and deletions, even when mutations occur much less frequently.

Taken together, our results show that the mutational dynamics observed in yeast are beneficial for the long-term stability of the genome and pave the way for a theory of evolution where genetic operators are themselves cause and outcome of the evolutionary dynamics.

Key words: genome evolution, ribosomal RNA gene locus, in silico evolution.

Introduction

In both bacteria and eukaryotes, genes that must be actively transcribed throughout a cell's life time may become the stage for clashes between transcription and replication complexes (Kim and Jinks-Robertson 2012; Merrikh et al. 2012; Brambati et al. 2014). Because these clashes can destabilize the genome and the epigenome (Takeuchi et al. 2003; Castellano-Pozo et al. 2013; Herrera-Moyano et al. 2014), several strategies have evolved to correctly schedule transcription and replication, for example, by slowing down the replication fork (Labib and Hodgson 2007), by releasing the mRNA faster (Bermejo et al. 2011) and in some cases by removing transcription complexes and degrading the nascent mRNA altogether (Brambati et al. 2014). Despite this, conflicts do occur and often result in mutations, ranging from single nucleotide substitutions to duplications and deletion due to recombination (Takeuchi et al. 2003; Kim and Jinks-Robertson 2012; Sankar et al. 2016).

One of the best characterized example of transcription–replication conflicts comes from the ribosomal RNA genes of *Saccharomyces cerevisiae*. Ribosomal RNA genes (rDNA) in yeast are organized in a cluster of about 150 copies on the XII chromosome. Large variations in rRNA gene copy number occur frequently in yeast (Ganley and Kobayashi 2011) and do not significantly alter growth rate—that is, they have no short-term fitness effect (French et al. 2003). The expression of rDNA genes is, in fact, independent of copy number and is instead under control of a dedicated polymerase (PolI) (French et al. 2003). When an rRNA gene is transcribed during DNA replication, collisions between transcription and replication complexes are avoided by a replication-fork barrier (RFB) (Takeuchi et al. 2003). RFBs are protein–DNA complexes located outside the coding region of each rRNA gene copy and stall the DNA polymerase coming from the opposite direction to transcription, preventing

head-on clashes between the Poll and the replication fork (Kobayashi et al. 1992). Prolonged RFB-mediated stalling can cause DNA polymerases to detach from the DNA, leaving a double-strand break behind with the break point corresponding to the genomic location of the RFB (Takeuchi et al. 2003; Krawczyk et al. 2014). These double-strand breaks happens more frequently in cells with fewer rDNA copies, because of larger transcriptional load on each gene (Kim et al. 2007; Ide et al. 2010). The break is repaired by some form of recombination, which can result in duplications and deletions (Ganley and Kobayashi 2011). Large rRNA transcriptional load can be triggered in yeast by placing cells under caloric excess after removing several rRNA gene copies. The Target Of Rapamycin (TOR) pathway (a ubiquitous nutrient signaling pathway) seems to control the rate of double-strand break formation, and thus that of duplications as well (Jack et al. 2015). Duplications prevail in these conditions, neutrally restoring the wild-type rDNA copy number distribution after few generations (Kobayashi et al. 1998; Jack et al. 2015). Therefore, yeast can control the rate of double-strand break formation, and thus that of duplications, depending on the availability of resources (by means of the TOR pathway, a ubiquitous nutrient signaling pathway) (Jack et al. 2015).

Overall, yeast does not adapt to different caloric conditions by selecting mutants with a precise number of rDNA repeats. Rather, it exploits a nutrient-dependent mechanism and an evolved genome structure to induce frequent but neutral mutations (gene duplications and deletions). These mutational dynamics rapidly restore the rDNA repeats when the copy number is small. Hence, yeast has evolved a genome structure that enhances the rate at which (part of) the genome evolves. In this article, we show that this evolution is adaptive, thereby showing a prime example of evolution of evolvability (Crombach and Hogeweg 2008; Hindré et al. 2012).

In order to understand the functional significance of the observed mutational regime in yeast, we study an in silico evolutionary model which integrates metabolism and transcription-induced mutagenesis. We find that the mutational biases observed empirically—in particular large rates of gene duplications and deletions—readily evolve in our model and are crucial to avoid the long-term accumulation of deleterious mutations. Thus, a bias in the types of mutations reduce mutational load despite increasing overall mutation rates.

Materials and Methods

We developed an in silico system to simulate the evolutionary dynamics of single-cell asexual populations (of constant size $N = 1{,}000$ individuals). Individual cells grow and divide, and integrate several features of the intracellular and genome dynamics of yeast (fig. 1).

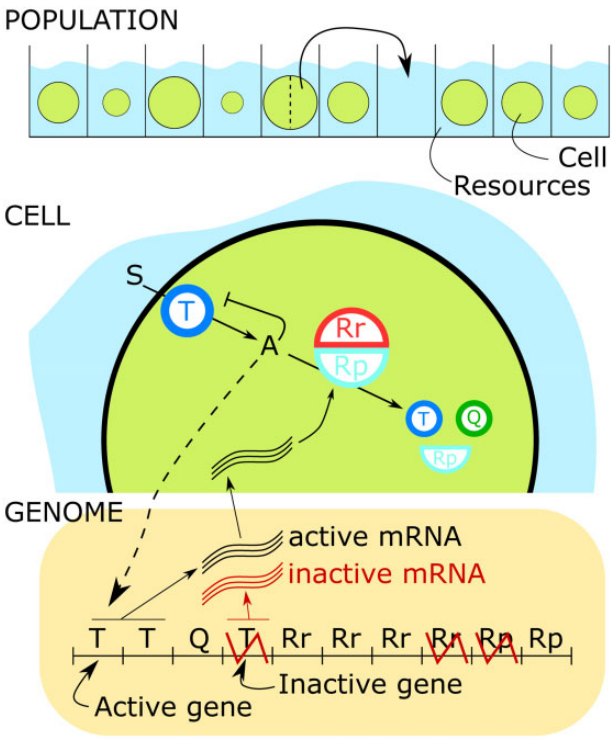

Fig. 1.—A cartoon of the model: an evolving population of yeastlike cells that grow and divide. Cell dynamics are formulated in terms of four macromolecules: enzymes T which convert resources S to amino acids A, housekeeping proteins Q which determine cell's health, ribosomal RNA Rr and ribosomal proteins Rp, which together translate mRNA into proteins. mRNA is transcribed from the cell's genome, which is composed of four gene types (possibly in multiple copies), one gene type per macromolecule. Genes can be duplicated, deleted, or inactivated by point mutations. See Materials and Methods and supplementary section S1, Supplementary Material online, for details.

Cell Dynamics

Each cell has a genome, a proteome and is capable of regulation. The genome is composed of four types of genes, one for each macromolecule: enzymes (T), housekeeping proteins (Q), ribosomal proteins (Rp), and ribosomal RNA (Rr). Each gene type can be present in multiple copies in the genome. Cell metabolism is loosely inspired by Scott et al. (2010): cells take up resources (S) from the environment and enzymes (T) convert them into amino acids (A). Genes are transcribed (mRNA is produced) with rates depending on a linear function of the amino acids (i.e., rate $= \max[0, k_0 + k_1 A]$, with regulation parameters k_0—the basal transcription rate—and k_1—the conversion factor for the amino acids) (Jewell et al. 2013) and homeostasis (i.e., a target concentration of housekeeping proteins Q). We do not model nucleotide metabolism, thus assuming that they are limiting for ribosome synthesis and cell growth. Per-gene transcription rates are capped at a maximum value k_{max} to model maximum packing of RNA polymerases over a gene and assumed to be the same value for all

genes. Ribosomal proteins (Rp) and RNAs (Rr) together form ribosomes, which convert amino acids into proteins based on the availability of mRNAs. Cells divide when target volume is reached. Target volume V_t is modeled as an increasing function of genome size G: $V_t = \kappa G^{\lambda}$ (with $\lambda = 0.9$).

Evolutionary Dynamics

We let fitness—the probability of replication—depends on interdivision time and homeostasis of housekeeping proteins. Fast growing cells (with smaller genomes) divide more frequently, and mutations occur upon cell division. Two categories of mutational events affect the genome: "background mutations," that is, random mutations independent of the cell life-history occurring with a constant (per gene) probability, and transcription-induced mutations (TIMs) caused by conflicts between transcription and replication complexes, occurring on highly transcribed genes. The latter are modeled as a steep sigmoid function of the transcription rates (averaged over a cell's interdivision time). This function ensures that these mutations occur rarely with low transcription rates (as is the case when transcription and replication can be correctly scheduled), and only when transcription rates are large (close to their maximum value) do conflicts—and therewith mutations—become frequent. Mutations result in gene duplications, deletions, and inactivations (point mutations) that cause irreversible loss of function. Inactivated genes maintain their promoter region and are therefore still transcribed but not translated. Unless otherwise stated, background duplications, deletions, and inactivations are always set to equal values ($\mu_{BG}^{dup} = \mu_{BG}^{del} = \mu_{BG}^{in} = 1/3 \times 10^{-4}$ per gene, per generation). Because transcription is assumed to be under polymerase control, copy number variation does not cause immediate dosage effect. This is in accordance with the transcription dynamics of rRNA genes (French et al. 2003), and it is extended to all genes in this model for consistency. Mutations of the regulation parameters (k_0, k_1) can occur during replication. In yeast rDNA, the effect of TIMs can be biased by the location of RFBs (at the end of each repeat). In our model, we study this mechanism by letting the relative proportion of duplications, deletions, and inactivations evolve. The relative rates of background mutations are fixed.

See supplementary section S1, Supplementary Material online, for all the details of the model and its main assumptions. Relevant parameter values are specified where needed, for all others see supplementary section S2, Supplementary Material online. All simulations used to generate the results presented in the Results section and supplementary material, Supplementary Material online, are listed in supplementary section S3, Supplementary Material online. The software used to run the model is available as supplementary material, Supplementary Material online.

Results

In the following, we first present our main result: Larger rates of duplications and deletions caused by large transcriptional load are beneficial to the long-term evolution of cells. Next, we connect this result to the mutational dynamics observed in the ribosomal DNA gene cluster of yeast. Finally, we show that an evolutionary feedback between regulation and gene copy number leads to genome deterioration when mutations are not biased toward duplications and deletions, and we explain why faster duplication–deletion dynamics prevent this deleterious evolutionary feedback.

Evolution Biases Transcriptional Mutagenesis toward Duplications and Deletions and Prevents Genome Deterioration

In accordance with experimental evidence, we assume that large transcriptional load is associated with higher mutation rates, and that evolution may bias the frequency of mutations toward duplications and deletions (e.g., by locating a RFB at the end of each gene in the cluster). Thus, the relative proportions of duplications, deletions, and inactivations associated with transcription can also evolve. In figure 2, we compare a population evolved without such mutations—that is, only with background mutations—to one with background mutations and evolvable proportions of TIMs (BG on the left in each pane, and BG + Ev.Tr. on the right). We let each system reach evolutionary steady state and collect data along the ancestral lineage (i.e., the lineage of common ancestors of the population) afterward.

We find that genomes at evolutionary steady state contain a large number of inactive genes in the case with only background mutations. In contrast, cells evolved with TIMs have a smaller steady-state genome because they do not accumulate many inactive genes, whereas the number of active genes is slightly larger and their interdivision time is shorter (see supplementary section S4, Supplementary Material online, for the evolutionary dynamics toward steady state). They also reach a fitter evolutionary steady state than those evolved with only background mutations. These cells maintain a large transcriptional load and therewith a high frequency of mutations (on average $\mu_{TR} = 0.065$, with the distribution skewed toward the maximum attainable value ≈ 0.074, as shown in figure 2 [Avrg.μ_{TR}]). We observe that the inactivation rate is minimized, whereas duplication and deletion rates are kept at the same proportions: The evolved (per gene) rates of duplications and deletions are ≈ 0.0325 each. In comparison, background mutations are set to 0.000033 for each type of mutation, about three orders of magnitude smaller than the evolved duplication and deletion rates. This result is surprising

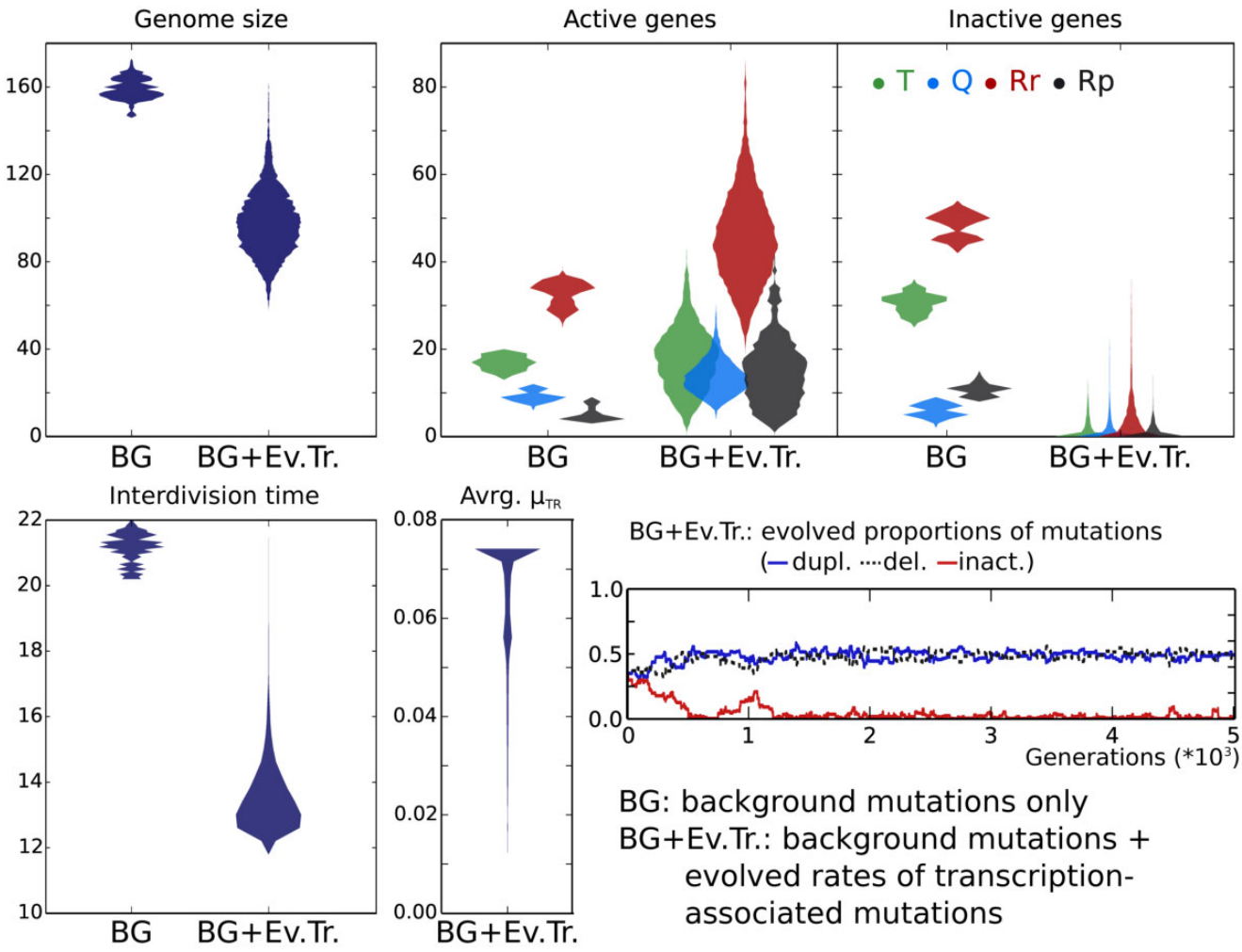

Fig. 2.—TIMs are beneficial when biased to duplications and deletions. Number of active and inactive genes: T, enzymes; Q, housekeeping proteins; Rr, Ribosomal RNA; Rp, ribosomal proteins. For both BG and BG + Ev.TR., background mutations $\mu_{BG} = 1/3 \times 10^{-4}$ for each type of mutation (duplications, deletions, and inactivations). Avrg. μ_{TR} is the TIM rate (per gene per replication), averaged over the four gene types. The maximum total mutation rate caused by transcription is $\max(\mu_{TR}) = 0.074$. Data collected along the ancestral lineage for 2,500 generations after evolutionary steady state is reached, that is, after $>10^6$ iterations (corresponding to an average of $>10^4$ generations).

because larger mutation rate leads to lower mutational load, thus avoiding the evolutionary deterioration of the genome.

Deletion of rRNA Genes Has Limited Effect on Fitness

Figure 3 shows that cell growth rate does not change as long as the rRNA gene copy number is sufficiently large, and it rapidly decreases when a large fraction of genes have been removed. Moreover, few generations are sufficient to generate large copy number variation in the ancestral lineage without much fitness difference (colored lines in fig. 3). We conclude that there is no direct benefit in the high copy number of rRNA genes in the evolved genomes. This is in qualitative agreement with yeast rDNA genome dynamics, where 1) high variability in rDNA copy number is observed (Ganley and Kobayashi 2011) and 2) removing copies of ribosomal genes does not affect short-term fitness (growth rate) but removing too many leads to a rapid deterioration of genome stability and, consequently, fitness (Ide et al. 2010).

A Larger Bias toward Duplications and Deletions Is Beneficial

To see how mutational biases lead to a fitter evolutionary steady state, we fix the ratio between transcription-induced duplications, deletions, and inactivations (whereas before it could evolve), progressively increasing the bias toward duplications and deletions. Figure 4 shows that a larger bias toward duplications and deletions contributes to a more compact genome and a shorter interdivision time. Populations evolved with only transcription-induced duplications and deletions (TR: large bias in fig. 4) have more active genes, and hardly

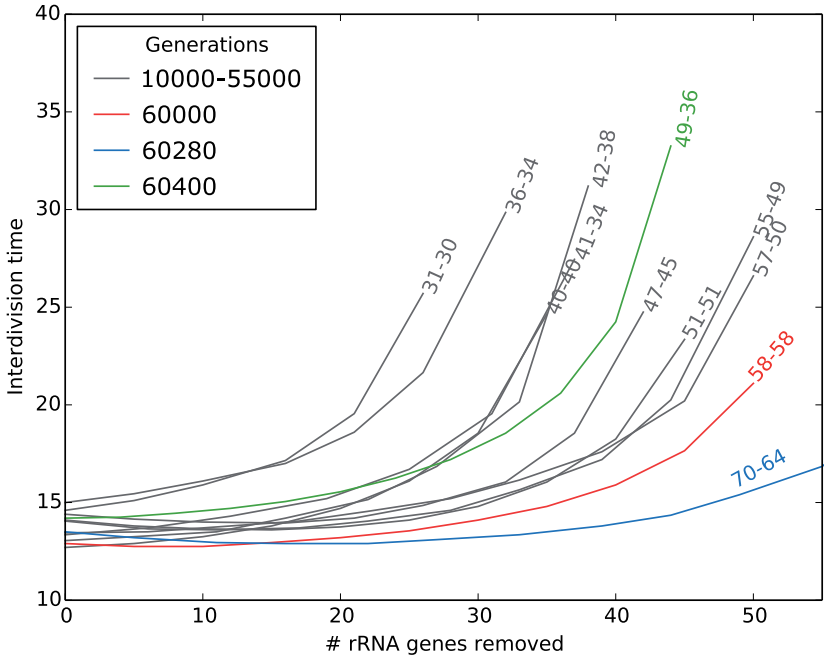

Fig. 3.—Growth rate is minimally affected when several ribosomal RNA genes are removed (cf., French et al. 2003). Genomes are taken from the same data set as figure 2 BG + Ev.Tr., along the ancestral lineage. Ribosomal RNA genes are removed maintaining the proportion between active and inactive ones and interdivision time is recorded. The original total and active rRNA gene copy number in each genome is indicated at the end of each plot. Colored lines highlight that few generations are sufficient for producing large variability in copy number.

any inactive genes. These cells maintain a large transcriptional load and consequently a high rate of per-gene duplications and deletions, similar to those presented in figure 2 BG + Ev.Tr. Conversely, cells with no bias in TIMs accumulate a large genome with more inactive genes and have much lower fitness. Moreover, these cells decrease their transcriptional load, thus reducing the frequency of TIMs.

A consequence of a steady and large transcriptional load is that mutations happen at a similar (large) rate every generation. This suggests a "conceptual" simplification of the model by introducing this bias directly in background mutations (i.e., disregarding that biased mutations are induced by transcription). The simplified model leads to similar results as above (supplementary section S5, Supplementary Material online). As a side note, if transcription-induced duplications occur more frequently than deletions, genomes accumulate inactive genes without bound, even in the absence of transcription-induced inactivations (see supplementary section S6, Supplementary Material online). Furthermore, background mutations do not have to happen at equal rates in our system (as they are likely not in yeast): TIMs can evolve to counterbalance preexisting biases in background mutations (supplementary section S7, Supplementary Material online). Finally, although we modeled the rate of TIMs as a sigmoid function that increases sharply when transcriptional load increases (see supplementary section S1, Supplementary Material online), our results are robust to alternative functions (supplementary section S8, Supplementary Material online).

Beyond Biasing Mutations: Large Rates of Duplications and Deletions Evolve

In the results presented so far, we assumed that a certain rate of TIMs is unavoidable. Therefore, the proportions of transcription-induced duplications, deletions, and inactivations were allowed to evolve, but their maximum total rate (per gene, per mRNA unit) was held constant. We next test whether evolution is solely selecting for lower inactivation rates, and the increase of duplications and deletions is a side effect of a constant total mutation rate. We relax the assumption of a maximum total mutation rate by letting the rates of transcription-induced duplications, deletions, and inactivations mutate independently from each other. In figure 5, we observe that larger rates of duplications and deletions evolve when background mutations occur frequently (we set $\mu_{BG}^{dup} = \mu_{BG}^{del} = \mu_{BG}^{in} = 0.0176$), while transcription-induced inactivations are quickly minimized. The evolved mutational dynamics lead to a fit evolutionary steady state (similar to fig. 2, despite background mutation rate being larger here). In yeast, evolution might only be able to alter the relative frequency at which mutations happen, for instance by increasing duplication–deletion dynamics by means of RFBs

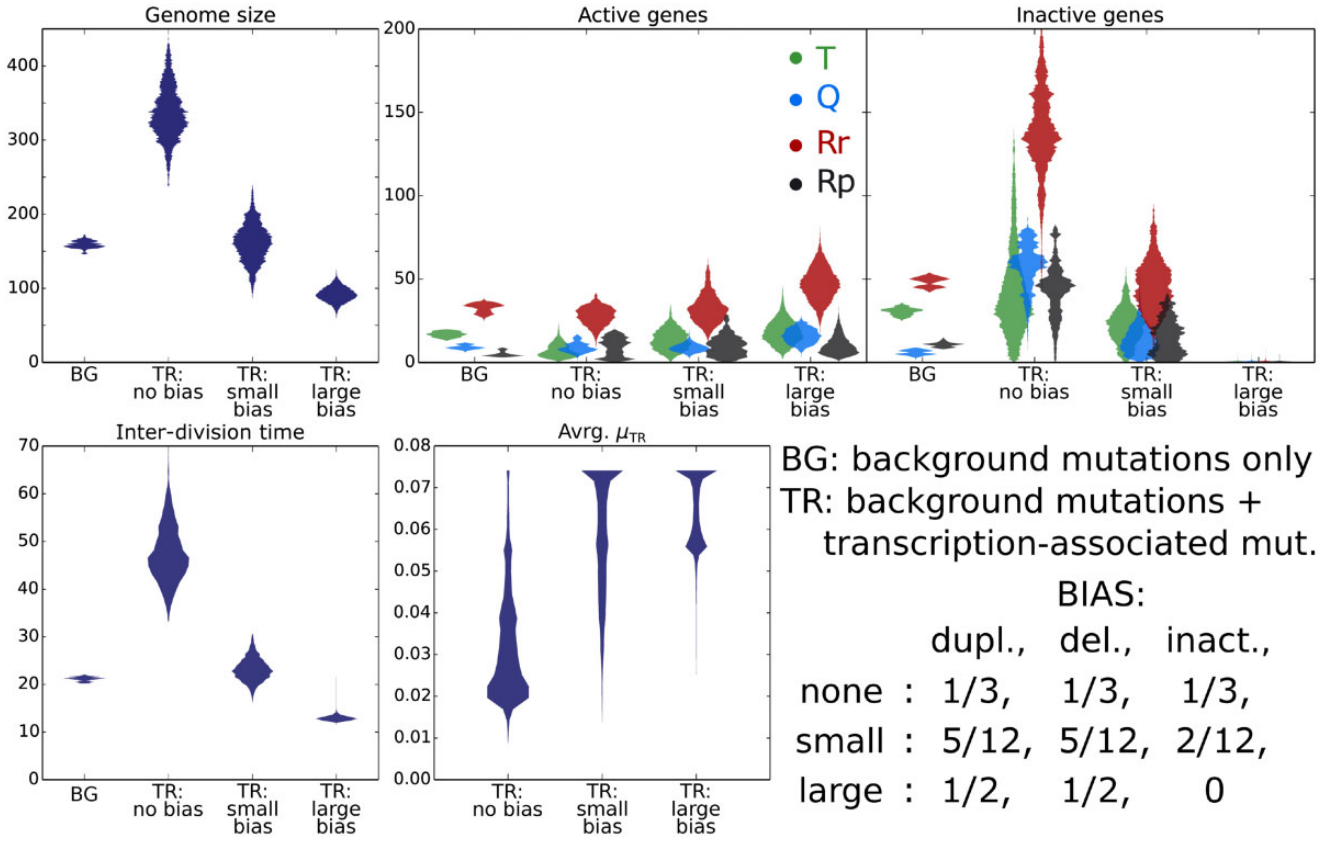

Fig. 4.—The beneficial effect of larger rates of TIMs, when biased to duplications and deletions. Data collected along the ancestral lineage for 2,500 generations after evolutionary steady state is reached. Background mutations (duplications, deletions, and inactivations) are equal in all four cases: $\mu_{BG} = 1/3 \times 10^{-4}$. The maximum mutation rate caused by transcription is $\max(\mu_{TR}) = 0.074$, the proportion of each mutation (i.e., the bias) is indicated in the figure.

(as shown in fig. 2, this strategy leads to high fitness). Letting the rates of duplications, deletions, and inactivations mutate independently might therefore not be directly applicable to yeast rRNA genes. Beyond modeling yeast, these results show that larger rates of duplications and deletions may be beneficial in rapidly evolving gene clusters.

Genome Deterioration Results from a Deleterious Evolutionary Feedback

The inability to purge inactive genes from the genome when mutations are not biased is reminiscent of Muller's ratchet (Muller 1964): the irreversible accumulation of slightly deleterious mutations due to drift in asexual populations. Muller's ratchet occurs in our system when a population of (initially) fit cells evolves with only background inactivating mutations (supplementary section S9, Supplementary Material online). Genomes deteriorate because mutations that inactivate single rRNA genes are selectively near neutral and can drift to fixation.

When duplications and deletions happen as frequently as inactivations, we observe that 1) active genes are maintained at a similar copy number with or without (unbiased) TIMs and 2) inactive genes accumulate (more dramatically in the case with TIMs; cf., fig. 4 TR: no bias and BG). This suggests that, after inactivating mutations, active genes are duplicated (to maintain their number) and inactive genes are inefficiently deleted. The effect is more pronounced for ribosomal RNA genes than for the other gene types. Because rDNA transcripts are not translated, more gene copies must be present in the genome to satisfy the metabolic requirements of a cell. This, in turn, intensifies the effects of the mutational dynamics and leads to the accumulation of a larger number of inactive genes in the long term.

In addition to the inefficient removal of inactive genes, a deleterious feedback establishes between the evolving regulation parameters and inactive genome inflation, when mutations are not biased. Because inactive genes are transcribed but not translated, transcriptional output effectively decreases after inactivations, and therewith growth rate. Because inactive genes are not efficiently purged from the genome, regulation parameters can adapt to increase transcription and can therefore compensate for a lower transcriptional output. Larger transcriptional demand leads to more TIMs, thereby

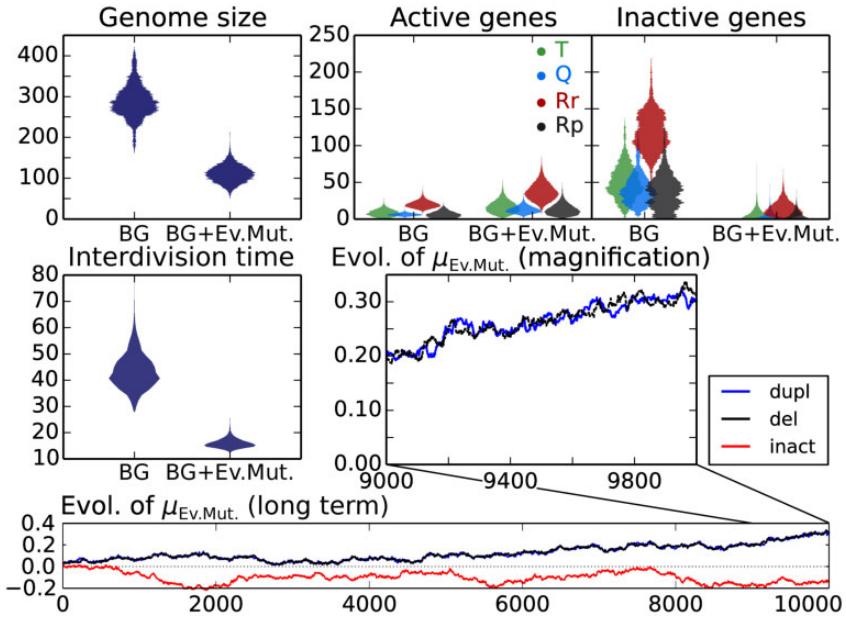

Fig. 5.—The evolution of TIMs. Background duplications, deletions, and inactivations occur at the same and equal rate $\mu_{BG} = 0.01763$ (per gene per replication) event both in the case without TIMs (BG) and with them (BG + Ev.Mut.). TIM rates $\mu_{Ev.Mut.}$ (per gene, per replication event) are allowed to evolve and may take values lower than zero, which is interpreted as zero within the simulation system. Mutation rates mutate (at replication) by a small random number from a uniform distribution with limits $[-\delta/2, \delta/2]$, with $\delta = 0.01$. Data collected along the ancestral lineage. Data for genome size and composition, and interdivision time are collected after reaching evolutionary steady state.

increasing the rate at which inactive genes accumulate. To show that evolved transcription rates lead to a larger inactive genome, we extract the regulation parameters (k_0 and k_1 for each gene type) from a simulation with equal rates of duplication, deletion, and inactivation (the same as shown in TR: no bias in fig. 4) at two time points: at early evolutionary time (50 generations) and at later time (20,000 generations). At the "early" evolutionary time, the deleterious evolutionary feedback between inactive genome and regulation parameters has not unfolded yet. The "late" time is after evolutionary steady state has been reached, and inactive genes have accumulated. We use these parameters to run two simulations in which the parameters themselves do not evolve (the mutation rates of regulation parameters k_0 and k_1 are set to zero). Inactive genes do not accumulate in the long run when "early" parameters are used, whereas they do with "late" parameters (fig. 6). Interdivision time mirrors this (notice that the number of coding genes is similar in the two cases) and, consistently, we observe that frequency distribution of mutations due to large transcriptional load is more spread with "late" parameters. This confirms that the evolutionary feedback between regulation parameters and genome composition leads to the accumulation of inactive (deleterious) genes.

High transcription rates are maintained when mutations are biased toward duplications and deletions (in fig. 4, Avrg. μ_{TR} is maintained close to the maximum, therefore transcription rates must be large). Furthermore, the "late" parameters achieve high transcription rates (fig. 6 late), as explained in the previous paragraph. From this, we expect that mutations biased to larger rates of duplications and deletions restore fitness with the "late" regulation parameters. This is confirmed in figure 6 (small and large bias) and demonstrates that despite the deleterious evolutionary feedback explained in the previous paragraph, the mutational dynamics (i.e., the relative proportions of duplications, deletions, and inactivations) ultimately determine genome composition at evolutionary steady state (at least in our model).

Large Rates of Duplications and Deletions Protect against Inactivations

So far, we have shown that frequent duplications and deletions overcome the deleterious accumulations of inactive genes. This is surprising in the context of Muller's ratchet, where higher mutation rates lead to increasing the rate of fixation of deleterious mutations. In this paragraph, we clarify how fitness is maintained despite higher mutation rates. We do this by setting up simplified simulations in which active and inactive gene dynamics are considered separately; for clarity, we allow only background mutations, that is, we do not allow TIMs and mutations of the regulation parameters. Results for the dynamics of rRNA genes are presented in figure 7. The other genes follow qualitatively the same dynamics, see supplementary section S10, Supplementary Material online, for the complete figure.

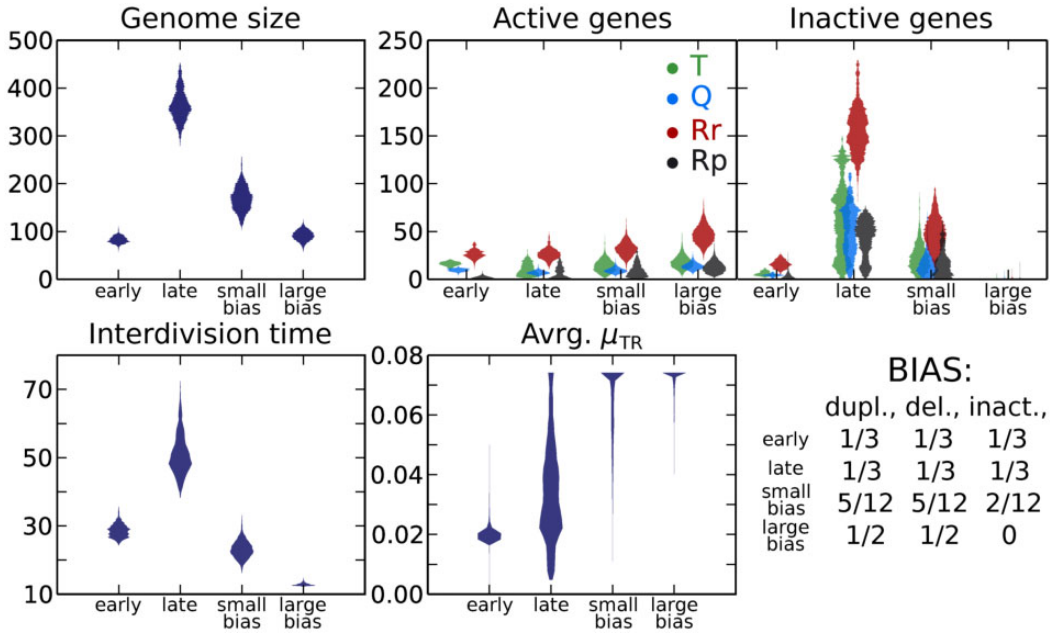

Fig. 6.—Disentangling the deleterious evolutionary feedback between regulation and genome composition: Mutational load increases with fixed regulation parameters previously evolved under equal rates of TIMs (cf., "early" vs. "late") and decreases with larger mutational bias toward duplications and deletions (cf., "late," "small bias," and "large bias"). Cells at an "early" and at a "late" stage of evolution with unbiased mutations (parameters as TR: no bias in figure 4 are used to seed a homogeneous population which cannot mutate its regulation parameters (other mutational dynamics are as before). In "small bias" and "large bias," the same regulation parameters as "late" are used, but mutations are more frequently duplications and deletions. After long-term evolution, data for the last 50 generations are collected for all runs to generate the figure. Values of "early" regulation parameters k_0, k_1: enzymes (0.099, 0.014), housekeeping proteins (0.11, 0.0078), ribosomal RNA (0.12, 0.031), and ribosomal proteins (0.11, 0.025). Values of "late" parameters (used also for "small bias" and "large bias"): enzyme (0.54, 0.39), housekeeping proteins (−0.24, 0.33), ribosomal RNA (0.13, 0.69), and ribosomal proteins (0.14, 0.21). Note that transcription rates depend on both regulation parameters and model dynamics.

We observe that high and low duplication and deletion rates lead to similar evolutionary steady states in all parameters when inactivations do not occur (fig. 7 active-Low and -High). Furthermore, in both cases high fitness is reached, as can be seen by comparing steady-state distributions to a control simulation with biased mutations (fig. 7 control-High). We conclude that selection for active gene copy number is not hindered by high rates of duplications and deletions. This is because no directional flow is imposed in genome space by duplications and deletions (if $\mu_{dup} = \mu_{del}$). Similarly, duplications can quickly restore the fittest copy number of active genes after inactivations occur, which explains why the number of active genes at evolutionary steady state does not differ drastically when different degrees of mutational bias are introduced (cf., active genes distributions in figure 7 control-Low, which is evolved with low mutation rates and no bias, to active and to fig. 4). Altogether, the system does not suffer from Muller's ratchet when duplications and deletions are the only mutations considered, regardless of their rate.

Because the steady-state number of active genes is stable, the number of genes inactivated per generation is approximately constant. We can use this to analyze the dynamics of inactive genes independently from those of active genes: We let duplications and deletions affect only inactive genes, and we let inactivations increase the number of inactive genes without decreasing active ones. We observe that low rates of duplications and deletions do not counter the inflow of inactive genes due to inactivations (fig. 7 inactive-Low, compare with control-Low). In contrast, high rates of duplications and deletions maintain low inactive gene copy number and therewith high fitness (fig. 7 inactive-High, cf., control-High).

To summarize, inactivations depend on μ_{in} and on the number of active genes, which is large and approximately constant no matter the rate of duplications and deletions. The deletion of inactive genes depends on μ_{del} and on their copy number (which is small, close to the fitness peak). In order to decrease the inactive genes, the number of deleted inactive genes per generation must be larger than the number of inactivations (given that fitness differences are small close to the fitness peak). To achieve this, μ_{del} must be sufficiently large (as fig. 7 inactive-High shows).

We conclude that frequent deletions are not deleterious for the maintenance of active genes but beneficial for reducing inactive ones, and larger rates of duplications do not affect the selection against inactive genes but restore the number of active genes after inactivations. Therefore, larger (and equal) rates of duplications and deletions prevent the accumulation of inactive genes.

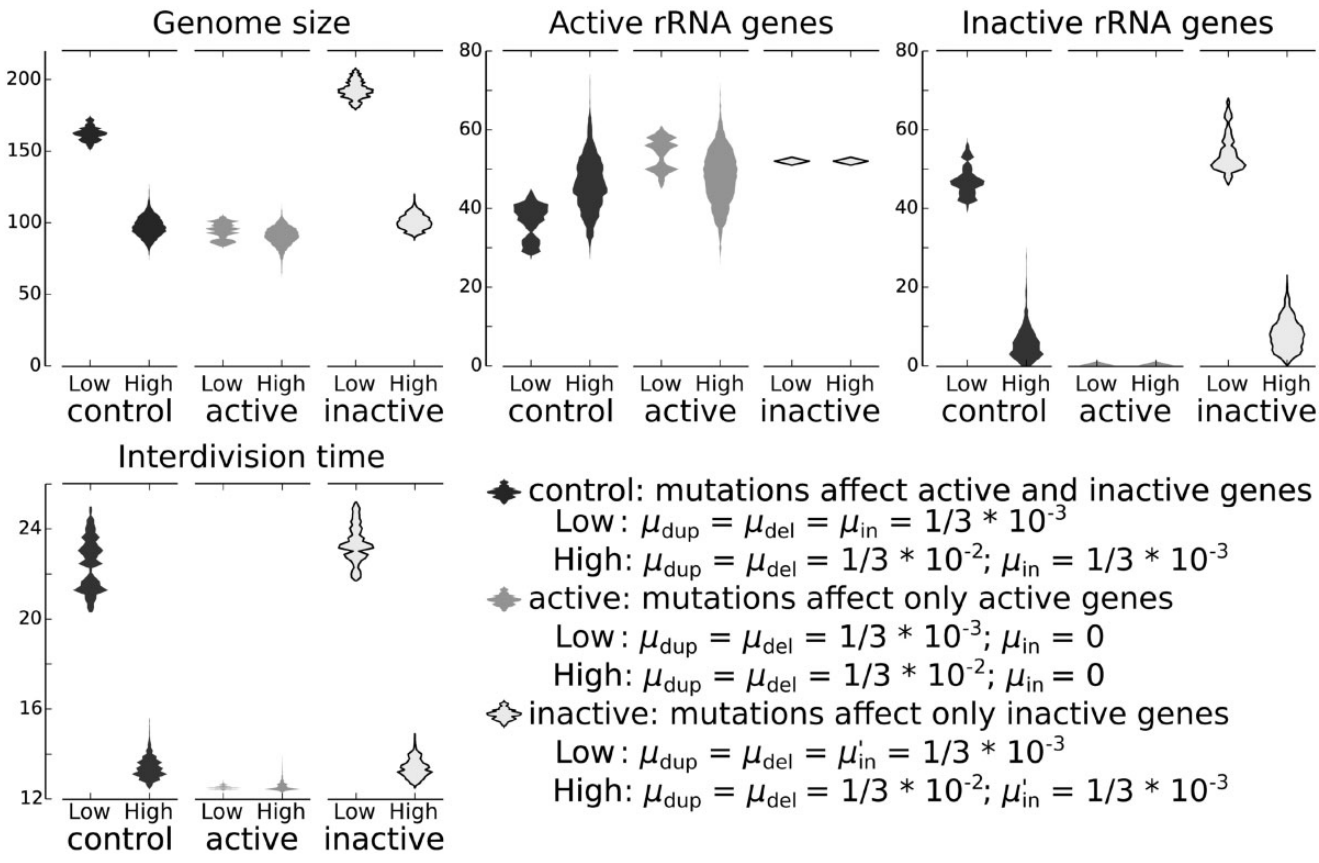

Fig. 7.—Large rates of duplications and deletions prevent the accumulation of inactive ribosomal RNA genes and do not affect selection for fittest number of active ones. Only background mutations are allowed. Only ribosomal RNA gene copy number is displayed in the active and inactive genes panes (see supplementary section S10, Supplementary Material online, for the other gene types). Note that for the case "inactive," new inactive genes are generated at rate μ'_{in} but the number of active genes is kept constant (as explained in the main text). Mutation rates are per gene, per replication. Initial populations for all runs are seeded with the same fit cell, chosen from the ancestral lineage of the same data set used for figure 2 BG + Ev.Tr., and stripping it of its (few) inactive genes so that its genome is entirely active.

Discussion

In this article, we studied the evolutionary dynamics of genomes in single-cell organisms by constructing a computational model in which metabolism, regulation, and mutational dynamics are integrated. Several theoretical models have been developed to study genome dynamics (Crombach and Hogeweg 2008; Cuypers and Hogeweg 2012), some at a greater resolution than what we do here (Knibbe et al. 2007). However, the effects of large transcriptional load on genome dynamics have received no attention (to the best of our knowledge), despite being extensively studied experimentally. Here, we made a first step in this direction.

We find that cells quickly adapt to grow rapidly (interdivision time being the fitness criterion), but their long-term evolution depends strongly on the mutational dynamics. Genomes accumulate inactive genes in the long evolutionary term when gene duplications, deletions, and inactivations are equally likely, because duplications of active genes are selected in the short run, but inactivations cannot be effectively purged. The accumulation of inactive genes caused by failure of selection to purge them is reminiscent of Muller's ratchet (Muller 1964). The surprising result of this article is that higher mutation rates can overcome it.

Genome integrity (as well as fitness) is maintained when transcription-induced duplications and deletions occur more frequently than inactivations, even when the overall mutation rate is increased. We have shown that frequent duplications and deletions do not decrease selectability of active gene copy number, and that higher duplication rates efficiently restore active gene copy number, whereas higher rates of deletions prevent the accumulation of inactive genes. These mutational dynamics suppress the evolutionary feedback that leads to the deterioration of the genome in the long term. The evolution of a mechanism to enhance copy number variation to maintain fitness (the evolution of the way in which the system evolves) is a form of evolution of evolvability. Note that the short-term effect of duplications and deletions is not the direct cause of increased fitness, as these mutations are almost

neutral (in contrast to, e.g., Crombach and Hogeweg [2008]). Their effect lies in overcoming the flow to inactivations by accelerating gene dynamics.

A similar situation is observed in yeast, where no short-term benefit is apparent despite the frequent occurrence of duplications and deletions in the rRNA gene cluster (French et al. 2003). In this respect, our work suggests that the observed mutational dynamics in yeast are beneficial for long-term genome integrity. Our model also shows that rDNA is prone to evolutionary deterioration, and depending on the actual rate of rRNA gene duplication and deletion, some accumulation of inactive genes might be inevitable. In support of this prediction, a recent bioinformatic analysis of the human genome showed that a considerable fraction of "junk" DNA is of ribosomal DNA origin (Robicheau et al. 2017). Furthermore, the large neutral variability in yeast rDNA copy number can be partly explained as an evolutionary side effect of the mutational dynamics. Additionally, yeast can silence the transcription of many copies of functional rRNA genes (Dammann et al. 1993). This redundancy enhances neutral copy number variability. We do not incorporate a silencing mechanism in our model, therefore underestimating the actual degree of variability in yeast rDNA.

Few mutational operators are defined in our model because we coarse-grained genomes to the gene level. As a consequence, genomes have little freedom in the coding structure they can evolve. Although this study shows that the effects of the mutational dynamics observed in yeast are beneficial, future work will address the evolution of the mechanisms of the regulated transcription-induced mutagenesis observed in yeast by incorporating finer details of genomes, such as evolvable fragile sites.

In our model, gene duplication or deletion does not immediately cause dosage effect and thus does not lead to large fitness differences. Similarly, inactivating mutations decrease transcription rates (and therewith metabolic rates) only slightly. Mutations have small fitness effects because we have assumed that transcription rates are not directly related to gene copy number. Hence, this assumption is at the root of the observed (Muller's ratchetlike) accumulation of inactive genes. Muller's ratchet is often discussed in the context of sexual reproduction. We are currently working on comparing the effect of duplications and deletions to that of recombination between different chromosomes, that is, comparing mutations due to asexual versus sexual reproduction. The lack of dosage effects is likely realistic only for rDNA among the genes in our model, but we extended it to all gene types. As a consequence, these genes behave as gene clusters (i.e., genes in high copy number), for which high rates of transcription-induced duplications and deletions are beneficial. One example of such gene cluster in the yeast genome (other than rDNA) is the copper resistance gene (CUP1) which is highly transcribed when yeast is under copper stress. Similarly to rDNA, TIMs in the CUP1 genes cause frequent duplications and deletions, and accelerate adaptation (Hull et al. 2017).

Nucleotide metabolism was not modeled in this study, thus assuming that it is limiting for ribosome production and consequently metabolism. Because nucleotides are "wasted" in the transcription of inactive ribosomal genes, nucleotide scarcity could select for shorter genomes. If this selection was strong enough to overcome the expansion of inactive genes, we would still expect a largely inactive genome at steady state (but shorter than what we found in this study). Nevertheless, higher rates of transcription-induced duplications and deletions would still be beneficial, because a compact and active genome does not waste nucleotides. Furthermore, we have implemented a minimal model of cell's metabolism and regulation. Amino acids are used as signaling molecules in our model, in accordance with their prominent role in the signaling cascade in the TOR pathway (Kim 2009; Jewell et al. 2013). However, the evolutionary consequences of richer regulation-mutation dynamics, such as those observed in yeast, remain to be explored.

In conclusion, conflicts between replication and transcription complexes may lead to mutations, especially for genes actively transcribed throughout the life cycle of a cell. The short-term benefit of duplications of active genes can lead to long-term deleterious consequences, that is, the accumulation of inactive genes, even at low mutation rates. Evolution can solve this problem by exploiting the increase in TIMs, to bias the mutational outcome toward duplications and deletions. We propose that this selection pressure generated the nutrient-dependent, RFB-mediated copy number variation observed in *Saccharomyces cerevisiae* and many other eukaryotic species (Eickbush and Eickbush 2007). The present study shows that these evolved mutational dynamics are beneficial in large evolutionary timescales, because they can prevent long-term genome deterioration.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

The authors thank N. Takeuchi and R.M.A. Vroomans for their help with conceptualizing this work. This work was partly supported by the European Commission 7th Framework Program (FP7-ICT-2013.9.6 FET Proactive: Evolving Living Technologies) EvoEvo project (ICT-610427). E.S.C. is supported by the Origins Center (NWA StartImpuls). The authors declare no competing interests.

Author Contributions

E.S.C. and P.H. designed the study, E.S.C. performed research, and E.S.C. and P.H. analyzed results and wrote the manuscript. Both authors have read and approved the final manuscript.

Literature Cited

Bermejo R, et al. 2011. The replication checkpoint protects fork stability by releasing transcribed genes from nuclear pores. Cell 146(2):233–246.

Brambati A, Colosio A, Zardoni L, Galanti L, Liberi G. 2014. Replication and transcription on a collision course: eukaryotic regulation mechanisms and implications for DNA stability. Front Genet. 6:166–166.

Castellano-Pozo M, et al. 2013. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. Mol Cell 52(4):583–590.

Crombach A, Hogeweg P. 2008. Evolution of evolvability in gene regulatory networks. PLoS Comput Biol. 4(7):e1000112.

Cuypers TD, Hogeweg P. 2012. Virtual genomes in flux: an interplay of neutrality and adaptability explains genome expansion and streamlining. Genome Biol Evol. 4(3):212–229.

Dammann R, Lucchini R, Koller T, Sogo JM. 1993. Chromatin structures and transcription of rDNA in yeast *Saccharomyces cerevisiae*. Nucleic Acids Res. 21(10):2331–2338.

Eickbush TH, Eickbush DG. 2007. Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics 175(2):477–485.

French SL, Osheim YN, Cioci F, Nomura M, Beyer AL. 2003. In exponentially growing *Saccharomyces cerevisiae* cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes. Mol Cell Biol. 23(5):1558–1568.

Ganley AR, Kobayashi T. 2011. Monitoring the rate and dynamics of concerted evolution in the ribosomal DNA repeats of *Saccharomyces cerevisiae* using experimental evolution. Mol Biol Evol. 28(10):2883–2891.

Herrera-Moyano E, Mergui X, García-Rubio ML, Barroso S, Aguilera A. 2014. The yeast and human fact chromatin-reorganizing complexes solve r-loop-mediated transcription–replication conflicts. Genes Dev. 28(7):735–748.

Hindré T, Knibbe C, Beslon G, Schneider D. 2012. New insights into bacterial adaptation through in vivo and in silico experimental evolution. Nat Rev Microbiol. 10(5):352–365.

Hull RM, Cruz C, Jack CV, Houseley J. 2017. Environmental change drives accelerated adaptation through stimulated copy number variation. PLoS Biol. 15(6):e2001333.

Ide S, Miyazaki T, Maki H, Kobayashi T. 2010. Abundance of ribosomal RNA gene copies maintains genome integrity. Science 327(5966):693–696.

Jack CV, et al. 2015. Regulation of ribosomal DNA amplification by the TOR pathway. Proc Natl Acad Sci U S A. 112(31):9674–9679.

Jewell JL, Russell RC, Guan K-L. 2013. Amino acid signalling upstream of mTOR. Nat Rev Mol Cell Biol. 14(3):133–139.

Kim E. 2009. Mechanisms of amino acid sensing in mTOR signaling pathway. Nutr Res Pract. 3(1):64–71.

Kim N, Abdulovic AL, Gealy R, Lippert MJ, Jinks-Robertson S. 2007. Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of DNA replication. DNA Repair 6(9):1285–1296.

Kim N, Jinks-Robertson S. 2012. Transcription as a source of genome instability. Nat Rev Genet. 13(3):204–214.

Knibbe C, Coulon A, Mazet O, Fayard J-M, Beslon G. 2007. A long-term evolutionary pressure on the amount of noncoding DNA. Mol Biol Evol. 24(10):2344–2353.

Kobayashi T, Heck DJ, Nomura M, Horiuchi T. 1998. Expansion and contraction of ribosomal DNA repeats in *Saccharomyces cerevisiae*: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. Genes Dev. 12(24):3821–3830.

Kobayashi T, Hidaka M, Nishizawa M, Horiuchi T. 1992. Identification of a site required for DNA replication fork blocking activity in the rRNA gene cluster in *Saccharomyces cerevisiae*. Mol Gen Genet. 233(3):355–362.

Krawczyk C, Dion V, Schär P, Fritsch O. 2014. Reversible top1 cleavage complexes are stabilized strand-specifically at the ribosomal replication fork barrier and contribute to ribosomal DNA stability. Nucleic Acids Res. 42(8):4985–4995.

Labib K, Hodgson B. 2007. Replication fork barriers: pausing for a break or stalling for time? EMBO Rep. 8(4):346–353.

Merrikh H, Zhang Y, Grossman AD, Wang JD. 2012. Replication–transcription conflicts in bacteria. Nat Rev Microbiol. 10(7):449–458.

Muller HJ. 1964. The relation of recombination to mutational advance. Mutat Res. 1(1):2–9.

Robicheau BM, Susko E, Harrigan AM, Snyder M. 2017. Ribosomal RNA genes contribute to the formation of pseudogenes and junk DNA in the human genome. Genome Biol Evol. 9(2):380–397.

Sankar TS, Wastuwidyaningtyas BD, Dong Y, Lewis SA, Wang JD. 2016. The nature of mutations induced by replication–transcription collisions. Nature 535(7610):178–181.

Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. 2010. Interdependence of cell growth and gene expression: origins and consequences. Science 330(6007):1099–1102.

Takeuchi Y, Horiuchi T, Kobayashi T. 2003. Transcription-dependent recombination and the role of fork collision in yeast rDNA. Genes Dev. 17(12):1497–1506.

Associate editor: Laurence Hurst