

Clonal interference between aneuploidy and mutation

Ilia Kohanovski¹ and Yoav Ram^{1,2,*}

¹School of Computer Sciences, IDC Herzliya, Herzliya, Israel

²School of Zoology, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

*Corresponding author: yoav@yoavram.com

June 6, 2021

Abstract

Aneuploidy is common in eukaryotes, often leading to decreased cell growth and fitness. However, evidence from yeast and fungi, as well as tumour cells, suggests that aneuploidy can be beneficial under stressful conditions and lead to elevated growth rates and to adaptation. Thus, it is crucial to develop a quantitative theory for the role of aneuploidy in adaptive evolution. Here, we analyse results from experiments in which *Saccharomyces cerevisiae* adapted to heat stress. The experimental population first acquired chromosome duplications, only to later revert back to a euploid state. We use several evolutionary models within an approximate Bayesian computation framework to show that clonal interference between chromosome duplications and beneficial mutations explains the experimental results. Our results suggest that aneuploidy can both accelerate and decelerate adaptation in a non-intuitive manner, creating an evolutionary conflict between the individual and the population and further complicating the process of adaptation in cell populations.

Introduction

Aneuploidy is common in eukaryotes. Aneuploidy is an imbalance in the number of chromosomes in the cell: an incorrect karyotype. Recent evidence suggests aneuploidy is very common in eukaryotes, e.g. animals (??), and fungi (????). Aneuploidy has been implicated in cancer formation and progression (??): 90% of solid tumours and 50% of blood cancers are aneuploid (?). Aneuploidy is also linked to the emergence of drug resistance (?) and virulence (?) in fungal pathogens, which are under-studied (?) despite infecting close to a billion people per year, causing serious infections and significant morbidity in >150 million people per year and killing >1.5 million people per year (?). In addition, aneuploidy is common in protozoan pathogens of the *Leishmania* genus, a major global health concern (?).

Aneuploidy is generally deleterious. The molecular and genetic mechanisms involved in aneuploidy have been explored in the past decade (?????). Experiments with human and mouse embryos found that aneuploidy is usually lethal; it is also associated with developmental defects and lethality in other multicellular organisms (?). For example, aneuploid mouse embryonic cells grow slower than euploid cells (?). Similarly, in unicellular eukaryotes growing in benign conditions, aneuploidy usually leads to slower growth and decreased overall fitness (?????), in part due to proteotoxic stress caused by increased expression in aneuploid cells (???) and hypo-osmotic-like stress (?).

Aneuploidy can lead to adaptation. However, aneuploidy can be beneficial under stressful conditions due to the wide range of phenotypes it can produce, some of which are advantageous (?). Thus, aneuploidy can lead to rapid adaptation in unicellular eukaryotes (????), as well as to rapid growth of somatic tumour cells (??). For example, aneuploidy in *S. cerevisiae* facilitates adaptation to a variety of stressful conditions like heat and pH (?), copper (??), salt (?), and nutrient limitation (??). Importantly, aneuploidy can also lead to drug resistance in pathogenic fungi such as *Candida albicans* (???) and *Cryptococcus neoformans* (?), which cause candidiasis and meningoencephalitis, respectively.

Transient adaptive solution. Aneuploidy differs from mutation due to its distinct properties. Chromosome duplication usually occurs more often than mutation and on average produces larger fitness effects. Yet, because it affects many genes on a whole chromosome or a chromosome fragment, aneuploidy also carries fitness costs. Thus, aneuploidy can be a *transient adaptive solution*: it can rapidly occur and fix in the population under stressful conditions, and can be rapidly lost when the cost outweighs the benefit – when stress is removed or after beneficial mutations occur. Experimental evidence of such a transient role of aneuploidy was demonstrated by ?, who evolved populations of *S. cerevisiae* under strong heat or pH stress. In these experiments, the populations quickly adapted to the stress, and this adaptation was determined to be due to chromosome duplications. Later on, the populations reverted back to an euploid state, while remaining adapted to the stress and accumulating multiple mutations. However, under gradual heat stress, aneuploidy was not observed. ? concluded that aneuploidy serves as a transient adaptive solution, or a “quick fix”, which is expected to facilitate adaptation.

The present study. Here, we develop a series of computational evolutionary models that include the effects of natural selection, genetic drift, aneuploidy and mutation to examine the role of aneuploidy in adaptation. These models follow a population of cells characterised by both their ploidy and their genotype. We fit these models to the empirical results of ? using approximate Bayesian computation. These model fits allow us to infer model parameters, including selection coefficients and transition rates, and to perform model selection between different models, thereby testing different hypotheses about the evolutionary process. We find that the aneuploidy rate is several orders of magnitude

higher than the mutation rate; that a simple model of clonal interference is not enough to explain the transience of aneuploidy; and that aneuploidy rescues the population, albeit at the price of delaying further adaptation.

Models and Methods

Evolutionary Models. To examine the role of aneuploidy in adaptation, two models were developed - *single-mutant model* and *multiple-locus model*. Both are based on the Wright-Fisher model (?) and include the effects of natural selection, genetic drift, aneuploidy, and mutation. Only beneficial mutations assumed, thus neglecting the effects of deleterious and neutral mutations. The *single-mutant model* allows for a single mutation (on specific locus or target loci), while the *multiple-locus model* allows for multiple mutations. Both models allow for a single aneuploid karyotype (e.g. chromosome III duplication). (Figure 1)

Single-mutant model. It contains four predefined genotypes whose frequencies change during generations. $2n$: euploid wildtype - the initial genotype; $2n^*$: euploid mutant, standard karyotype with one specific beneficial mutation; $2n+1$: aneuploid wildtype, with an extra chromosome, e.g. following chromosome duplication; $2n+1^*$: aneuploid mutant, with both the beneficial mutation and the extra chromosome.

The parameters of the model are: N , the population size, which is constant in all generations; μ , the mutation rate - the transition rate from $2n$ to $2n^*$, i.e. the frequency at which cells with genotype $2n^*$ are created from cells with genotype $2n$; α , the aneuploidy gain rate - the transition rate from $2n$ to $2n+1$; α_l , the aneuploidy loss rate - the transition rate from $2n+1^*$ to $2n^*$; k , the coefficient, such that $k \cdot \mu$ is the transition rate from $2n+1$ to $2n+1^*$, where on default k is equals to 33/32 (euploid $2n$ genotyped *S. cerevisiae* contains 32 chromosomes, therefore, assuming each chromosome have equal probability to have the specific beneficial mutation in, aneuploid $2n+1$ genotype should have 33/32 times the mutation rate); s , the mutation selection coefficient; s_b , fitness benefit of aneuploidy; c , fitness cost of aneuploidy, where $0 < c < 1$. The relative fitnesses of genotypes are:

$$\begin{aligned} w_{2n} &= 1 \\ w_{2n+1} &= 1 - c + s_b \\ w_{2n+1^*} &= (1 + s)(1 - c) + s_b \\ w_{2n^*} &= (1 + s) \end{aligned} \quad (1)$$

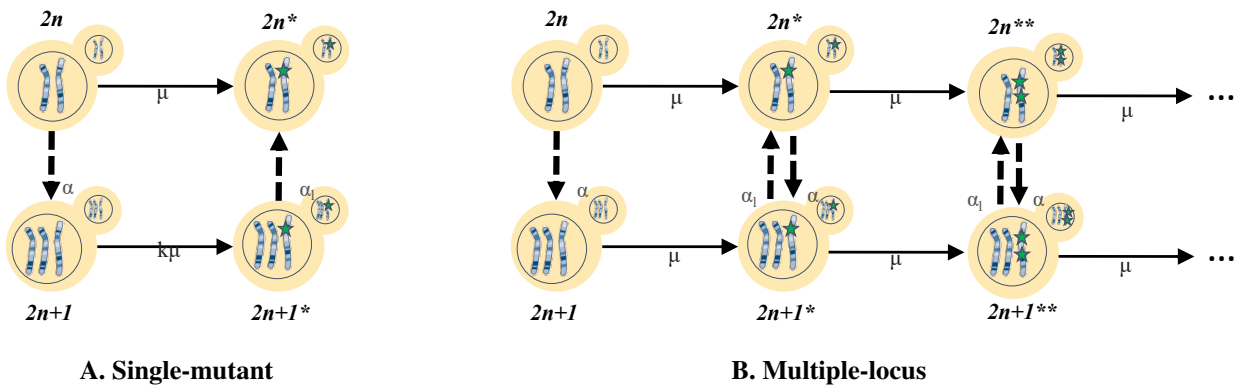


Figure 1: Model Schemes. (A) The population with four possible genotypes and four possible transitions from one genotype to another with the transition rates μ , α , α_l , $k\mu$. $2n$: euploid wildtype – the initial genotype; $2n^*$: euploid mutant, standard karyotype with one specific beneficial mutation; $2n+1$: aneuploid wildtype, with an extra chromosome, e.g. following chromosome duplication; $2n+1^*$: aneuploid mutant, with both the beneficial mutation and the extra chromosome. Overall there are two possible paths to reach $2n^*$ from $2n$. (B) Each genotype can have zero or more beneficial mutations and zero or one extra chromosomes. μ , mutation rate; α , aneuploidy gain rate; α_l , aneuploidy loss rate

The model follows Wright-Fisher (?) steps. The first generation is initialized with N cells of the genotype $2n$. At each generation, the frequency f_g of each genotype g is updated:

1. Selection step:

$$f_g \leftarrow \frac{f_g w_g}{\sum_j f_j w_j} \quad (2)$$

2. Mutation gain and aneuploidy gain/loss step:

$$\begin{aligned} f_{2n} &\leftarrow f_{2n} - \alpha f_{2n} - \mu f_{2n} \\ f_{2n+1} &\leftarrow f_{2n+1} - k\mu f_{2n+1} + \alpha f_{2n} \\ f_{2n+1*} &\leftarrow f_{2n+1*} - \alpha_l f_{2n+1*} + k\mu f_{2n+1} \\ f_{2n*} &\leftarrow \mu f_{2n} + \alpha_l f_{2n+1} \end{aligned} \quad (3)$$

3. Random drift step: draw the frequencies of all genotypes from a multinomial distribution

$$f \sim \text{Mult}(N, (f_{2n}, f_{2n+1}, f_{2n+1*}, f_{2n*})) \quad (4)$$

Multiple-locus model. It expands the *single-mutant model* by allowing for any number of beneficial mutations. For each mutation i , selection coefficient s_i is independently drawn from an exponential distribution with expected value s . The parameters of the model are μ , the mutation rate; α , the aneuploidy gain rate; α_l , the aneuploidy loss rate; s , the parameter of the exponential distribution; s_b , fitness benefit of aneuploidy; c , fitness cost of aneuploidy, where $0 < c < 1$. In this model, the population may include any number of genotypes that contain zero or more mutations and zero or one extra chromosomes. The fitness w_g of genotype g is:

$$w_g = \begin{cases} 1, & \text{wildtype} \\ 1 - c + s_b, & \text{aneuploid wildtype} \\ \prod_{i \text{ is mutation in } g} (1 + s_i), & \text{euploid mutant} \\ s_b + (1 - c) \prod_{i \text{ is mutation in } g} (1 + s_i), & \text{aneuploid mutant} \end{cases} \quad (5)$$

Therefore, aneuploidy loss would be favored by selection only if there are enough beneficial mutations or/and the mutation selection coefficients s_i are large enough. The intuition is as follows: when the benefit of the accumulated beneficial mutations is small, then the benefit of aneuploidy has a large effect; when the benefit of the accumulated beneficial mutations benefit is large, then aneuploidy doesn't add much, and its cost becomes significant.

Unlike the single-locus model, the population size for the multiple-locus model is not constant. Indeed, a serial-transfer protocol (?) is simulated, where the population is repeatedly diluted by transfer to a fresh medium, starting a new growth cycle, similar to the original experiment. In the model the population initial size is N , then it is doubled every iteration till reaching $8 \cdot N$, then it is diluted back to N .

Denote by n_g be the number of cells of genotype g in the population. At each generation, genotypes frequencies are changed by the following steps:

1. Selection step:

$$n_g \leftarrow n_g \cdot w_g \quad (6)$$

2. Random drift step:

$$f \sim \text{Mult}(N', (\frac{n_1}{N'}, \frac{n_2}{N'}, \dots)) \text{ where } N' = n_1 + n_2 + \dots \quad (7)$$

3. Normalization after population size N change (growth or dilution):

$$n_g \leftarrow f_g \cdot N \quad (8)$$

4. Mutation step:

For each genotype g , draw the number of mutant cells from a binomial distribution $Bin(n_g, \mu)$, and subtract that number from n_g . For each mutant cell, generate a new genotype and add one cell of this genotype to the population. The newly created genotype inherits mutations and ploidy karyotype from the parent genotype g , and in addition gains one more mutation, with the selected coefficient drawn from the exponential distribution $Exp(\lambda)$. The fitness of the newly created genotype is assigned according to the Formula 5.

5. Aneuploidy gain step:

For each genotype g that doesn't already have an aneuploid chromosome, draw the number of new cells that gain aneuploidy from a binomial distribution $z \sim Bin(n_g, \alpha)$. If $z > 0$, then subtract it from n_g and create a new genotype and add to the population with z cells of this genotype. The genotype inherits all the mutations of genotype g and in addition gains aneuploidy. The fitness is calculated appropriately by the Formula 5.

6. Aneuploidy loss step:

It is similar to the aneuploidy step. For each genotype g that has aneuploidy, draw the number of new cells that lose aneuploidy from the binomial distribution $z \sim Bin(n_g, \alpha_l)$. If $z > 0$, subtract z from n_g , then create new genotype, and add to the population z cells of this genotype. The genotype inherits all the mutations of g and in addition gains aneuploidy. The fitness is calculated appropriately by the Formula 5.

Data. We inference parameters of the models based on the data of the experiments of Yona et al. In *heat-stress* experiment four populations of *S. cerevisiae* were evolved under 39° heat-stress and developed aneuploidy till 450 generations (here and further by development/loss of genotype trait we mean that 95% of the population are the carriers of the genotype). Then, two populations were checked at generations 1700 and 2350, and for one population the aneuploidy were eliminated and has not observed already at generation 1700 and for the second at generation 2350. Similarly, in *ph-stress* experiment four populations were evolved under high-ph stress conditions and developed aneuploidy during 150 generations. At generation 750 no aneuploidy was observed for all four populations.

Likelihood function. Denote by A_{450} the event of aneuploidy observation at generation 450; by L_{1700} the event of aneuploidy loss observation at generation 1700; and by L_{2350^*} the event of aneuploidy loss observation at generation 2350 while there's no loss in generation 1700. The likelihood function of the model with parameter set ϕ for *heat-stress* experiment is:

$$L(\phi | A_{450}, L_{1700}, L_{2350^*}) = P(L_{450})^4 \cdot (1 - P(\neg L_{1700} | A_{450})^4 - P(\neg L_{2350^*} | A_{450})^4 + P(\neg L_{1700} \wedge \neg L_{2350^*} | A_{450})^4) \quad (9)$$

Results

Discussion

Aneuploidy and mutation: *same-same but different*. The published data indicate that, like mutation, aneuploidy can be both deleterious and beneficial (??). Nevertheless, there are important and fundamental differences between adaptation by aneuploidy and adaptation by beneficial mutations (?), which make aneuploidy a unique mechanism for generating genetic variation. First, the aneuploidy rate (i.e. the frequency of mis-segregation events) is significantly higher than the mutation rate (?). Thus, everything else being equal, adaptation by aneuploidy will be faster and more frequent. Second, fitness effects of aneuploidy are larger than those of the majority of mutations, on average, and are rarely neutral (???), allowing selection to quickly sort deleterious and beneficial genotypes. Third, the number of different karyotypes is considerably smaller than the number of different genotypes, and different karyotypes are likely to have different phenotypes (?). Therefore, exploration of the phenotype space by aneuploidy requires smaller populations and a shorter time span. Fourth, aneuploidy is a reversible state, as the rate of chromosome loss is high and the cost of aneuploidy is significant (?). Indeed, aneuploidy often provides a transient solution: under short-term stress conditions, aneuploidy reverts (chromosome number returns to normal) when the stress subsides; under long-term stress conditions, aneuploidy reverts when refined solutions, generated by beneficial mutations, take over (?). Finally, aneuploidy results in increased genome instability, potentially increasing genetic variation by a positive feedback loop (???), while also increasing its own transience.

Evolutionary theory of aneuploidy. The role of aneuploidy in adaptation has only recently been observed (???), and is largely missing from the literature on evolution and adaptation: the introductory textbook *Evolution* by ? does not mention the word aneuploidy, and the graduate-level book *Mutation-Driven Evolution* by ? only briefly mentions aneuploidy in the context of speciation, but not adaptation. In recent reviews of the literature, aneuploidy is suggested to play an important role in fungal adaptation (??) and cancer evolution (???), yet these reviews cite no theoretical studies nor any quantitative models. Indeed, evolutionary, ecological, and epidemiological studies mostly assume adaptation occurs via beneficial mutations, recombination, and sex. Therefore, there is a critical need to develop an evolutionary theory of aneuploidy like the evolutionary theories of other mechanisms for generation of genetic variation, e.g. mutation (?), recombination (?), and sex (?). An evolutionary theory of aneuploidy will be central to the interpretation of experimental and clinical observations and design of new hypotheses, experiments, and treatments (?). For example, despite the lack of theoretical models, aneuploidy has been invoked in a new strategy to combat pathogens and tumour cells by setting “evolutionary traps” (??), in which a condition that predictably leads to emergence of aneuploidy is applied, followed by a condition that specifically selects against aneuploid cells.

Acknowledgements

We thank Lilach Hadany, Judith Berman, David Gresham, and Martin Kupiec for discussions and comments. This work was supported in part by the Israel Science Foundation 552/19 and Minerva Stiftung Center for Lab Evolution (YR).