

Adaptive evolution with aneuploidy and mutation

Ilia Kohanovski^{1,*}, Martin Pontz^{2,*}, Avihu H. Yona³, 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

³Institute of Biochemistry, Food Science and Nutrition, Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Israel

*These authors contributed equally to this work

†Corresponding author: yoav@yoavram.com

December 15, 2021

Abstract

Aneuploidy is common in eukaryotes, often leading to decreased cell growth and fitness. However, evidence from yeast and fungi, as well as human tumour cells, suggests that aneuploidy can be beneficial under stressful conditions and lead to elevated growth rates and adaptation. Because aneuploidy differs from mutation in rate, expected effect, and reversibility, it is crucial to develop a quantitative theory for the role of aneuploidy in adaptive evolution. Here, we develop evolutionary models for adaptive evolution with both mutation and aneuploidy. These models are used within an approximate Bayesian computation framework to estimate the formation rate and fitness effect of aneuploidy and mutation from empirical results of 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 also analyze our models to estimate the effect of the aneuploidy and mutation rates on the expected adaptation time and the probability for adaptation via aneuploidy. Our results suggest that aneuploidy is a transient adaptive solution, which can decelerate adaptation in a non-intuitive manner. By creating an evolutionary conflict between the individual and the population, aneuploidy further complicates 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. Evidence suggests aneuploidy is very common in eukaryotes, e.g. animals (Santaguida and Amon, 2015; Naylor and van Deursen, 2016; Bakhoun and Landau, 2017), and fungi (Pavelka et al., 2010; Zhu et al., 2016; Robbins et al., 2017; Todd et al., 2017). Aneuploidy has been implicated in cancer formation and progression (Boveri, 2008; Schwartzman et al., 2010): 90% of solid tumours and 50% of blood cancers are aneuploid (Santaguida and Amon, 2015). Aneuploidy is also linked to the emergence of drug resistance (Selmecki et al., 2009) and virulence (Möller et al., 2018) in fungal pathogens, which are under-studied (Rodrigues and Albuquerque, 2018) 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 (Selmecki et al., 2009; Rodrigues and Albuquerque, 2018). In addition, aneuploidy is common in protozoan pathogens of the *Leishmania* genus, a major global health concern (Mannaert et al., 2012).

Aneuploidy is generally deleterious. The molecular and genetic mechanisms involved in aneuploidy have been explored (Musacchio and Salmon, 2007; Sheltzer and Amon, 2011; Chen et al., 2012; Rancati and Pavelka, 2013; Gerstein et al., 2015; Shor and Perlin, 2015). 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 (Sheltzer and Amon, 2011). For example, aneuploid mouse embryonic cells grow slower than euploid cells (Williams et al., 2008). Similarly, in unicellular eukaryotes growing in benign conditions, aneuploidy usually leads to slower growth and decreased overall fitness (Niwa et al., 2006; Torres et al., 2007; Pavelka et al., 2010; Sheltzer and Amon, 2011; Kasuga et al., 2016), in part due to proteotoxic stress caused by increased expression in aneuploid cells (Pavelka et al., 2010; Santaguida et al., 2015; Zhu et al., 2018) and hypo-osmotic-like stress (Tsai et al., 2019).

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 (Pavelka et al., 2010). Thus, aneuploidy can lead to rapid adaptation in unicellular eukaryotes (Gerstein et al., 2015; Torres et al., 2010; Hong and Gresham, 2014; Rancati et al., 2008), as well as to rapid growth of somatic tumour cells (Schwartzman et al., 2010; Sheltzer et al., 2017). For example, aneuploidy in *S. cerevisiae* facilitates adaptation to a variety of stressful conditions like heat and pH (Yona et al., 2012), copper (Covo et al., 2014; Gerstein et al., 2015), salt (Dhar et al., 2011), and nutrient limitation (Dunham et al., 2002; Gresham et al., 2008). Importantly, aneuploidy can also lead to drug

resistance in pathogenic fungi such as *Candida albicans* (Selmecki et al., 2008, 2010; Gerstein and Berman, 2018) and *Cryptococcus neoformans* (Sionov et al., 2010), 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 Yona et al. (2012). They evolved populations of *S. cerevisiae* under strong heat or pH stress. The populations adapted to the heat and pH stress within 450 and 150 generations, and this adaptation was determined to be due to chromosome duplications. Much later, after more than 1500 and 750 generations, for the heat and pH stress, respectively, 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. Yona et al. (2012) concluded that aneuploidy serves as a transient adaptive solution, or a “quick fix”, which is expected to facilitate adaptation.

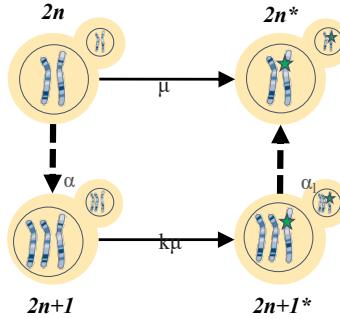
The present study. Here, we develop evolutionary-genetic models that include the effects of natural selection, genetic drift, aneuploidy, and mutation to examine the role of aneuploidy in adaptive evolution. These models follow a population of cells characterised by both their ploidy and their genotype. We fit these models to the experimental results of Yona et al. (2012) using an *approximate Bayesian computation* framework (Sisson et al., 2007; Klinger et al., 2018) to infer model parameters, including selection coefficients and rates of aneuploidy and mutation, and to perform model selection between different models, thereby testing different hypotheses about the evolutionary process. Furthermore, we analyze these models to estimate the effects of parameters on the adaptation time and the probability for adaptation via aneuploidy. We find that

Models and Methods

Evolutionary Models. We developed two models: a single-locus model and a multi-locus model. Both models are based on the Wright-Fisher model (Otto and Day, 2007), assuming non-overlapping generations and including the effects of natural selection, genetic drift, aneuploidy, and mutation. We focus on beneficial mutations, neglecting the effects of deleterious and neutral mutations. Both models allow for a single aneuploid karyotype (e.g., chromosome III duplication). While the single-locus model allows for only a single mutation to occur, the multi-locus model allows for multiple mutations to accumulate in the genome (Figure 1).

Single-locus model. This model assumes a constant population size N and follows four genotypes (Figure 1A): euploid wild-type, $2n$, the initial genotype; euploid mutant, $2n^*$, with the standard

A. Single-locus model



B. Multi-locus model

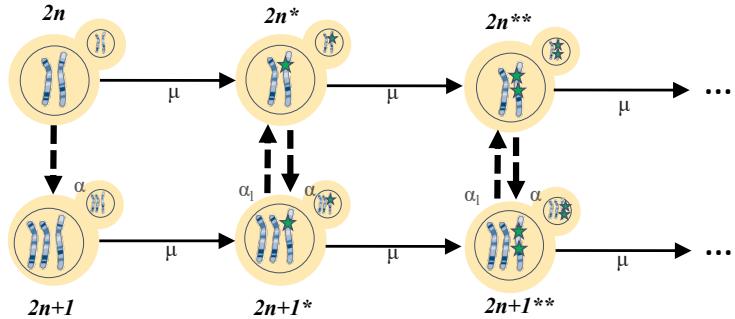


Figure 1: Model illustrations. **(A)** In the single-locus model, the four genotypes are: euploid wild-type, $2n$; euploid mutant, $2n^*$; aneuploid wild-type, $2n+I$; and aneuploid mutant, $2n+I^*$. Overall there are two possible trajectories from $2n$ to $2n^*$. **(B)** In the multi-locus model, each genotype is characterized by its karyotype, $2n$ or $2n+I$, and the number of accumulated beneficial mutations, denoted by stars. In both panels arrows denote transitions between genotypes, with transitions rates: μ , mutation rate; δ , aneuploidy gain rate; δ_L , aneuploidy loss rate.

karyotype and a single beneficial mutation; aneuploid wild-type, $2n+1$, with an extra chromosome, e.g. following chromosome duplication; and aneuploid mutant, $2n+1^*$, with an extra chromosome and a beneficial mutation.

Transitions between the genotypes occur as follows (Figure 1A): Beneficial mutations from $2n$ to $2n^*$ occur with probability μ , the mutation rate, and from $2n+1$ to $2n+1^*$ with probability $\tau\mu$, where τ is the fold-change in the rate of beneficial mutations in aneuploid cells. By default, we assume $\tau = 33/32$, as the wild-type *S. cerevisiae* strains in the experiments by (Yona et al., 2012) are diploid, with 32 chromosomes, and the aneuploid strains are trisomic, with 33 chromosomes. Aneuploidy is formed by chromosome mis-segregation, so that cells transition from $2n$ to $2n+1$ with probability δ , the aneuploidy gain rate. Aneuploidy is lost, transitioning cells from $2n+1^*$ to $2n^*$ with probability δ_L , the aneuploidy loss rate. The fitness values of the four genotypes are given by Table 1.

Table 1: Single-locus model fitness values.

Genotype i	$2n$	$2n + 1$	$2n + 1^*$	$2n^*$
Fitness w_i	1	$1 - c + b$	$(1 - c)(1 + s) + b$	$1 + s$

$s \geq 0$ is the selection coefficient of a beneficial mutation; $b \geq 0$ is the selection coefficient of aneuploidy; and $0 \leq c \leq 1$ is the fitness cost of aneuploidy.

The first generation is initialized with N cells with genotype $2n$. The effect of natural selection on the frequency f_i of genotype $i = 2n, 2n + 1, 2n + 1^*$, or $2n^*$ is given by

$$f_i^s = \frac{f_i w_i}{\bar{w}} , \quad (1)$$

where the fitness values w_i are given in Table 1 and $\bar{w} = \sum_j f_j w_j$ is the population mean fitness. The effect of mutation and aneuploidy on genotype frequencies is given by

$$\begin{aligned} f_{2n}^m &= (1 - \delta - \mu) f_{2n}^s , \\ f_{2n+1}^m &= \delta f_{2n}^s + (1 - \tau\mu) f_{2n+1}^s , \\ f_{2n+1^*}^m &= \tau\mu f_{2n+1}^s + (1 - \delta_L) f_{2n+1^*}^s , \\ f_{2n^*}^m &= \mu f_{2n}^s + \delta_L f_{2n+1}^s + f_{2n^*}^s . \end{aligned} \quad (2)$$

Finally, random genetic drift is modeled using a multinomial distribution (Otto and Day, 2007),

$$\mathbf{f}' \sim \frac{1}{N} Mult(N, \mathbf{f}^m) , \quad (3)$$

where $\mathbf{f}^m = (f_{2n}^m, f_{2n+1}^m, f_{2n+1^*}^m, f_{2n^*}^m)$ are the frequencies of the genotypes after mutation and aneuploidy, \mathbf{f}' are the genotype frequencies in the next generation, and $Mult(N, \mathbf{f})$ is a multinomial distribution

parameterized by the population size N and the genotype frequencies \mathbf{f} . Overall, the change in genotype frequencies from one generation to the next is given by the transformation $f_i \rightarrow f'_i$.

Multi-locus model. This model expands the single-locus model by allowing for (i) the accumulation of beneficial mutations in the genome, and (ii) a fluctuating population size.

A genotype is characterized by its karyotype, $2n$ or $2n+1$, and the number of accumulated beneficial mutations, which can be zero or more. The selection coefficient of the i -th accumulated mutation in each individual, s_i , is drawn from an exponential distribution with expected value s , $s_i \sim \text{Exp}(s)$. The rest of the parameters ($N, \mu, \tau, \delta, \delta_L, b, c$) are the same as in the single-locus model. The fitness of the different genotypes is the same as in the single-locus model (Table 1), except that the fitness contribution of k beneficial mutations is the product of their independent effects, $\prod_{i=1}^k (1 + s_i)$, instead of the contribution of the single mutation allowed in the single-locus model, $(1 + s)$, see Table 2. Therefore, aneuploidy loss would be favored by selection only if there are enough beneficial mutations and/or the selection coefficients s_i are large enough. The intuition is that when the benefit of the accumulated beneficial mutations is small, the benefit of aneuploidy has a large effect; when the benefit of the accumulated beneficial mutations benefit is large, then aneuploidy is no longer advantageous because of its significant cost.

In contrast to the single-locus model, in the multi-locus model the population size changes in order to model serial-transfer experiment protocol (Yona et al., 2012): the population is serially diluted by transferring a fraction of the population (1/120) to a fresh medium, starting a new growth cycle. In this model, the population initial size is $N_0 = N$, and the population size is doubled every generation, $N_1 = 2N, N_2 = 4N, \dots$, and diluted back to N every eight generations, $N_8 = N$.

The change in frequencies due to selection is exactly the same as in the single-locus model (Equation 1), only applied using the fitness values in Table 2. The change due to random genetic drift is also the same as in Equation 3, except that the frequencies vector is $\mathbf{f} = (f_{2n}, f_{2n+1}, f_{2n^*}, f_{2n+1^*}, f_{2n^{**}}, f_{2n+1^{**}}, \dots)$ and that the population size changes between generations, as described above.

The effects of mutation and aneuploidy on genotype frequencies is more elaborate than in the single-locus model. Genotype i is classified according to their karyotype ($2n$ or $2n+1$), the number of accumulated beneficial mutations ($k \geq 0$), and their fitness (w_i). Each offspring cell inherits these properties from its mother cell. Then, with probability μ or $\tau\mu$ for euploid and aneuploid cells, respectively, a new beneficial mutation is accumulated, such that the number of mutations is $k + 1$, and its effect s_{k+1} is drawn from an exponential distribution with expected value s , such that the

contribution of the mutations to the fitness is $\prod_{j=0}^{k+1} (1 + s_j)$. Next, euploid offspring become aneuploid with probability δ , and aneuploid offspring become euploid with probability δ_L .

Table 2: Multi-locus model fitness values.

<i>Genotype i</i>	$2n$	$2n + 1$	$2n + 1^{*k}$	$2n^{*k}$
<i>Fitness w_i</i>	1	$1 - c + b$	$(1 - c) \prod_{j=1}^k (1 + s_i) + b$	$\prod_{j=1}^k (1 + s_i)$

k is the number of accumulated beneficial mutations in the genome; $s \geq 0$ is the selection coefficient of a beneficial mutation; $b \geq 0$ is the selection coefficient of aneuploidy; and $0 \leq c \leq 1$ is the fitness cost of aneuploidy.

Empirical evidence. Our inference procedure uses empirical data from evolutionary experiments performed by Yona et al. (2012). In the heat-stress experiment, four populations of *S. cerevisiae* evolved under 39 °C. Aneuploidy fixed in all four population in the first 450 generations (hereafter, fixation or elimination of a genotype by generation t means that more than 95% or less than 5% of the population carry the genotype at generation t , and possibly earlier). By unpublished observations the aneuploidy was not fixed before at least 200 generations. The experiment continued with two populations, in which aneuploidy was eliminated by generation 1700 and 2350. In the pH-stress experiment, four populations of *S. cerevisiae* evolved under high-ph stress (8.6). Aneuploidy fixed during the first 150 generations. It was fully eliminated in two populations and partly eliminated in two populations by generation 750. These empirical results were published by Yona et al. (2012).

Likelihood function. Denote by $A_{t,T}$ the fixation of aneuploidy at generation t , conditioned on no fixation at generation T ; by L_t the loss of aneuploidy at generation t ; and by $L_{t,T}^*$ the loss of aneuploidy at generation t conditioned on no loss by generation T . The model likelihood function for parameter vector θ for the heat-stress experiment is

$$\begin{aligned} \mathcal{L}(\theta | A_{450,200}, L_{1700}, L_{2350,1700}^*) = \\ P^4(A_{450,200}) \cdot \left(1 - P^4(\neg L_{1700} | A_{450,200}) - P^4(\neg L_{2350,1700}^* | A_{450,200}) + P^4(\neg L_{1700} \wedge \neg L_{2350,1700}^* | A_{450,200}) \right), \end{aligned} \quad (4)$$

where $P(X)$ is the probability of event X and \neg means *not*.

The likelihood function of the model with parameter set θ for the pH-stress experiment is

$$\mathcal{L}(\theta | A_{150}, L_{750}) = P^4(L_{150}) \cdot 6 \cdot P^2(L_{600} | A_{150}) \cdot P^2(\neg L_{600} | A_{150}). \quad (5)$$

For the model without aneuploidy, let's denote by M_t , the fixation of $2n^*$ at generation t , and by $M_{t,T}^*$ the fixation of $2n^*$ at generation t conditioned on no fixation at generation T . The likelihood of the model without aneuploidy with parameter vector θ for the heat-stress experiment is

$$\begin{aligned}\mathcal{L}(\theta|M_{1700}, M_{2350,1700}^*) = \\ \left(1 - P^4(\neg M_{1700}) - P^4(\neg M_{2350,1700}^*) + P^4(\neg M_{1700} \wedge \neg M_{2350,1700}^*)\right), \quad (6)\end{aligned}$$

Because of the complexity of the model, these likelihood functions are intractable. Therefore, given specific values of the model parameters θ , we simulate the model in many replicates to approximate the value of the likelihood function $\mathcal{L}(\theta)$. For the single locus model 1000 replicates were used.

Parameter inference. To infer model parameters, we use approximate Bayesian computation (Sunåker et al., 2013) with a sequential Monte-Carlo scheme, or ABC-SMC (Sisson et al., 2007), implemented in the pyABC¹ Python package (Klinger et al., 2018). Briefly, this approach uses numerical stochastic simulations of the model to infer a posterior distribution over the model parameters. It is a method of likelihood-free, simulation-based inference (Cranmer et al., 2020), that is, for estimating a posterior distribution when a likelihood function cannot be computed. It is therefore suitable to our case, in which the likelihood function can only be approximated from simulations, and cannot be directly computed. In short, the algorithm employs sequential importance sampling over generations. In generation t , the algorithm draws sets of parameters θ_i from a given proposal distribution and consequently simulates data C_i from the model until n_t instances were accepted based on the comparison to observed data C_{obs} via a distance function $D(C_i, C_{obs})$ and acceptance threshold ϵ_t , such that $D(C_i, C_{obs}) \leq \epsilon_t$. As the distance function we choose $1 - \mathcal{L}$. For the acceptance threshold the median of the distances of the accepted instances of the previous generation are applied (for the first generation we define the acceptance threshold be 0.01). When n_t instances are accepted new proposal distribution is constructed by multivariate normal kernel with adaptive covariance matrix. In the first generation the prior (see below section) defines the proposal distribution. The size of n_t is determined by adaptive population strategy (Klinger and Hasenauer, 2017). We stop the iterations, when no more sufficient change in acceptance threshold occur, and no change in posterior. We stop single-locus model after four iterations with epsilon 0.13, when 1074 accepted instances with effective sample size of 700 (Figure S5). We ran the algorithm with different initialization seeds and different number of replicates for simulations and got similar posterior (Figure S6).

After getting weighted particles of the posterior distribution by ABC-SMC scheme, we calculate

¹<https://pyabc.readthedocs.io>

Kernel Density Estimate using Gaussian kernels and 50,000 points are sampled from KDE. We use the KDE and the points to calculate mode, highest density interval (HDI), for drawing distribution plots, etc.

Model comparison. We perform model selection using WAIC (widely applicable information criterion) (Gelman et al., 2013),

$$WAIC(\theta | \mathbf{X}) = -2 \log \mathbb{E}[\mathcal{L}(\theta | \mathbf{X})] + 2\mathbb{V}[\log \mathcal{L}(\theta | \mathbf{X})] \quad (7)$$

where $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ are the expectation and variance operators taken over the posterior distribution $P(\theta | \mathbf{X})$. WAIC values are scaled as a deviance measure: lower values imply higher predictive accuracy (Kass and Raftery, 1995).

Prior distributions. The prior distributions of $w_{2n+1} = 1 - c + b$, $w_{2n+1^*} = (1 + s)(1 - c) + b$ and $w_{2n^*} = 1 + s$ were obtained by estimating it from growth curves data that were previously obtained by Yona et al. (2012) using mono-culture growth experiments. The raw growth curves data were not published before, but were used to produce figures 3C, 4A, and S2 in Yona et al. (2012). We used Curveball², a dedicated method for predicting results of competition experiments from growth curve data (Ram et al., 2019). Curveball takes growth curves of two strains growing separately in mono-culture and predicts how they would grow in a mixed culture, that is, it predicts the results of a competition assay. From these predictions, relative fitness values can be computed.

We used growth curves of $2n+1$ and $2n$ in 39 °C to estimate w_{2n+1} . We used the same prior for w_{2n+1^*} and w_{2n^*} , assuming that fitness values shouldn't differ too much. We also tested different priors, but this one (denote it *main prior*) was better in the sense of WAIC, prediction plots, estimated parameters. We tested uniform prior (denote *uninformative*), $Uniform(1, 6)$, for (i) all w_{2n+1} , w_{2n+1^*} , w_{2n^*} and (ii) only for w_{2n+1^*} , w_{2n^*} , leaving w_{2n+1} prior as in main prior. In both scenarios the fitness estimates by the model looks too high: for the first scenario the median estimates were $w_{2n+1} = 2.767$, $w_{2n+1^*} = 5.79$, $w_{2n^*} = 5.827$ and for the second $w_{2n+1} = 1.046$, $w_{2n+1^*} = 4.642$, $w_{2n^*} = 4.684$. We don't expect to have such high fitness values. Moreover the WAIC for the models with these priors are 0.57 and 1.9 respectively, while WAIC for the main prior model is 0.27. Also for the model with uninformative prior for all w_{2n+1} , w_{2n+1^*} , w_{2n^*} the fixation time of $2n+1$ is about 16 in average, which contradicts unpublished observations that were at least 250-300.

Also we tried to use additional growth curves. We used growth curves of $2n^*$ (*refined* from Yona et al. (2012)) and $2n+1$ in 39 °C to estimate w_{2n^*}/w_{2n+1} . The same prior was used for w_{2n^*}/w_{2n+1^*} . This

²<https://curveball.yoavram.com>

model didn't perform well, it resulted in WAIC 2.8.

Another try was to use growth curves of $2n + 1$ and $2n$ in 30°C to estimate $1 - c$. This estimation assumes that the cost of aneuploidy is the same in 39°C and 30°C ; this might be incorrect, but we only assumed this to generate a prior distribution for the fitness values. The prior for b was taken the same as for c . This model didn't work at all: no parameter sets were found with the likelihood greater than zero.

Because Curveball uses a maximum-likelihood approach to estimate model parameters, we were able to estimate a distribution of relative fitness values by sampling from a truncated multivariate normal distribution, defined by the maximum-likelihood covariance matrix. Thus, we sampled 10,000 values for appropriate fitness, which we used as prior distributions. See Figures S1 and S2.

For the mutation rate μ and aneuploidy rate δ we used uniform priors, $\text{Uniform}(1e - 8, 1e - 5)$ and $\text{Uniform}(1e - 6, 1e - 3)$ respectively. We tried also uniform priors with lower than the $1e - 8$ bound for μ , but anyway the posterior ended with higher values. For the single locus model without aneuploidy, wider uniform prior for the mutation rate was used - $\text{Uniform}(1e - 10, 1e - 5)$.

Results

Model fitting

Single-locus model parameters inference. First we inference the parameters of the single-locus model. The mode and 50% hdi of the parameters posterior are: mutation rate - $3.06_{-2.84}^{+1.229} \times 10^{-6}$, trisomy rate - $1.61_{-0.197}^{+1.122} \times 10^{-3}$, $2n+1$ fitness - $1.022_{-0.002}^{+0.001}$, $2n+1^*$ fitness - $1.025_{-0.001}^{+0.001}$, $2n^*$ fitness - $1.028_{-0.002}^{+0.001}$,

This trisomy rate appropriates to the previous estimates; the mutation rate corresponds to the mutations target size of 10^4 if we assume the mutation rate of single base to be about $2 * 10^{-10}$ as estimated by Zhu et al. (2014). We can see from the posterior distribution (Figure 2) the correlation between fitness of different genotypes.. (TODO say/plot something about the correlation, is it interesting?).

Single-locus model fits the data well, but only if contains aneuploidy. Single-locus model fits well to the data (Figure S3A). On Figure 3 the dynamics of frequency change of each genotype in time is shown. We can see that $2n+1^*$ never reach substantial frequency in the population. Sensitivity analysis (Figure S4) shows that changing the parameter values cause the model to fit less. Furthermore, we measured fixation time of $2n+1$ in population, and found it appropriate to the unpublished data of the experiment (Yona et al., 2012) - about 300 generations in average, that supports the correctness

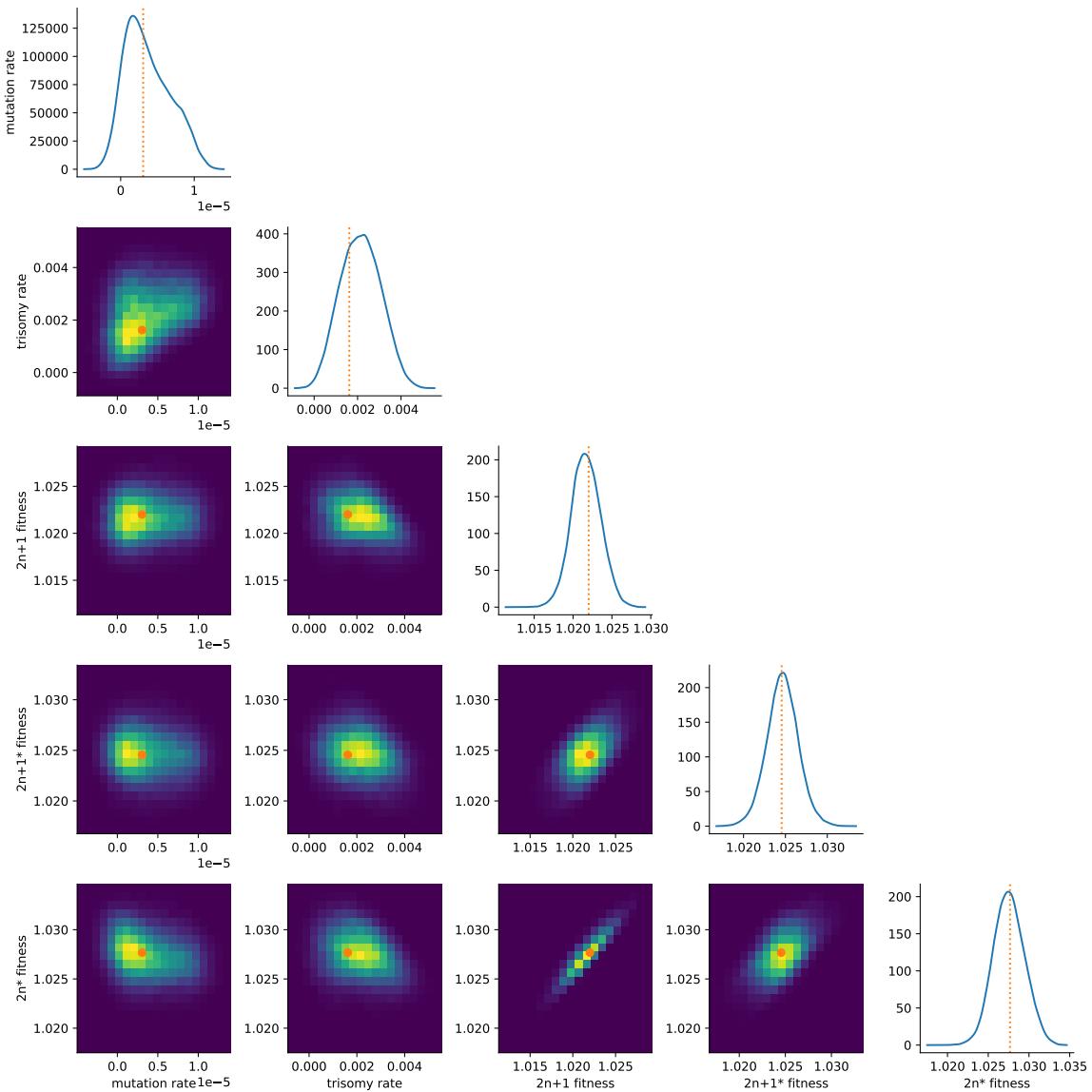


Figure 2: Single-locus model parameters posterior density. On the diagonal plots, the posterior density for each parameter is presented . Other plots represent joint posterior density of two parameters where blue color has the lowest density and yellow has the highest density. Red point and the dashed lines represent the maximum a posteriori probability estimate: mutation rate - 3.06×10^{-6} , trisomy rate - 1.61×10^{-4} , $2n+1$ fitness - 1.022, $2n+1^*$ fitness - 1.025, $2n^*$ fitness - 1.028.

of the inference. On the other hand, if we don't take into consideration aneuploidy in the model, i.e. take aneuploidy rate be equal to zero, then the model can't explain the data: the fit is not so good (Figure S3A; Figure 4). The inference parameter values of the model without aneuploidy are: mutation rate - $1.251_{-0.177}^{+0.146} \times 10^{-8}$, $2n+1$ fitness - $1.01_{-0.002}^{+0.001}$, $2n+1^*$ fitness - $1.011_{-0.001}^{+0.001}$, $2n^*$ fitness - $1.011_{-0.001}^{+0.001}$. We can see that mutation rate for this model is much lower than for the model with aneuploidy. Higher mutation rate would cause quicker fixation of $2n^*$ than expected by the experiment (Figure 4 blue bar). Thereby, adding aneuploidy to the model, helps to attain good fit with higher

mutation rate.

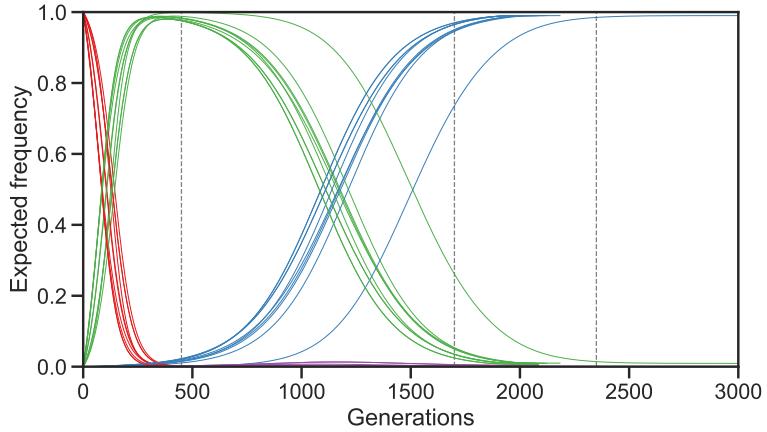


Figure 3: Evolutionary dynamics with fitted single-locus model. Frequency of the four genotypes averaged over 10,000 simulations of the model. Each line represent one of the 10 parameter sets drawn randomly from the inference posterior.

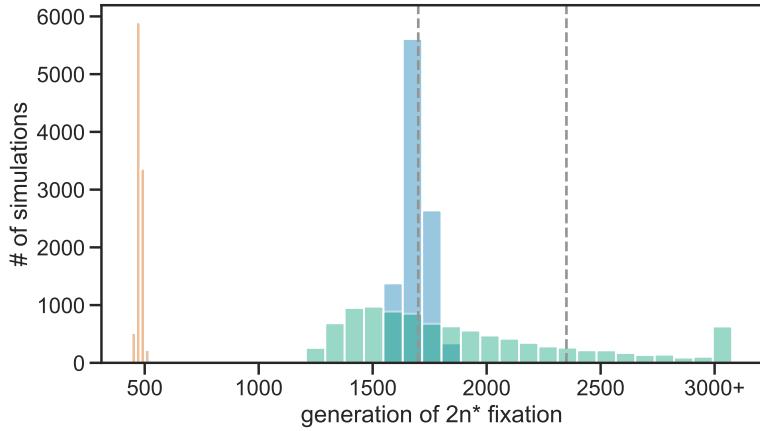


Figure 4: Time of $2n^*$ fixation cannot be explained without aneuploidy. Each color represents distribution to fixation of $2n^*$ of 10,000 simulations for different modifications of single-locus model. *Blue* bars correspond to the model with aneuploidy, where parameters are taken from the median of the posterior of the inference. *Orange* bars correspond to the same model with same parameters, except aneuploidy that is zero. *Green* bars correspond to the model without aneuploidy, i.e. with aneuploidy rate zero, where parameters are taken from the median of the posterior of the inference for this model. The last bin contains all the simulations with time equal or greater than 3000. Dashed lines represent experimental data: generations 1700 and 2350. When the difference in number of simulations between ranges 0-1700 and 1700-2350 is small, and there are no much simulations after generation 2350 - the fit is good. We can see that the fit of the model with aneuploidy (*blue*) is the best. There's no fit at all for the model that is represented by *yellow* bars.

Complex model: Inference and Comparison.

Discussion

Aneuploidy is not just another type of mutation. The published data indicate that, like mutation, aneuploidy can be both deleterious and beneficial (Pavelka et al., 2010; Sheltzer and Amon, 2011). Nevertheless, there are important and fundamental differences between adaptation by aneuploidy and adaptation by beneficial mutations (Yona et al., 2015), 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 (Santaguida and Amon, 2015). 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 (Pavelka et al., 2010; Yona et al., 2012; Sunshine et al., 2015), 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 (Pavelka et al., 2010). 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 (Niwa et al., 2006). 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 (Yona et al., 2012). Finally, aneuploidy results in increased genome instability, potentially increasing genetic variation by a positive feedback loop (Rancati and Pavelka, 2013; Bouchonville et al., 2009; Zhu et al., 2012), while also increasing its own transience.

Evolutionary theory of aneuploidy. The role of aneuploidy in adaptation has only recently been observed (Sionov et al., 2010; Yona et al., 2012; Gerstein et al., 2015), and is largely missing from the literature on evolution and adaptation: the introductory textbook *Evolution* by Bergstrom and Dugatkin (2012) does not mention the word aneuploidy, and the graduate-level book *Mutation-Driven Evolution* by Nei (2013) 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 (Robbins et al., 2017; Todd et al., 2017) and cancer evolution (Santaguida and Amon, 2015; Naylor and van Deursen, 2016; Sansregret and Swanton, 2017), 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 (Lynch, 2010), recombination (Hartfield and Keightley, 2012), and sex (Otto, 2009). An evolutionary theory of aneuploidy will be central to the interpretation of experimental and clinical observations and design of new hypotheses, experiments, and treatments (Carja et al., 2014). 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” (Gerstein et al., 2015; Chen et al., 2015), 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 Yitzhak Pilpel, Orna Dahan, Lilach Hadany, Judith Berman, David Gresham, Shay Covo, Martin Kupiec, and Tal Simon for discussions and comments. This work was supported in part by the Israel Science Foundation (YR 552/19) and Minerva Stiftung Center for Lab Evolution (YR).

References

- Bakhoum, S. F. and Landau, D. A. (2017), ‘Chromosomal instability as a driver of tumor heterogeneity and evolution’, *Cold Spring Harb. Perspect. Med.* **7**(6), 1–14.
- Bergstrom, C. T. and Dugatkin, L. A. (2012), *Evolution*, 1 edn, W. W. Norton & Company, New York, NY.
- Bouchonville, K., Forche, A., Tang, K. E. S., Semple, C. a. M. and Berman, J. (2009), ‘Aneuploid chromosomes are highly unstable during DNA transformation of *Candida albicans*.’, *Eukaryot. Cell* **8**(10), 1554–66.
- Boveri, T. (2008), ‘Concerning the Origin of Malignant Tumours’, *J. Cell Sci.* **121**(Supplement 1), 1–84.
- Carja, O., Liberman, U. and Feldman, M. W. (2014), ‘Evolution in changing environments: Modifiers of mutation, recombination, and migration’, *Proc. Natl. Acad. Sci.* p. 201417664.
- Chen, G., Mulla, W. A., Kucharavy, A., Tsai, H.-J., Rubinstein, B., Conkright, J., McCroskey, S., Bradford, W. D., Weems, L., Haug, J. S., Seidel, C. W., Berman, J. and Li, R. (2015), ‘Targeting the Adaptability of Heterogeneous Aneuploids’, *Cell* **160**(4), 771–784.
- Chen, G., Rubinstein, B. and Li, R. (2012), ‘Whole chromosome aneuploidy: Big mutations drive adaptation by phenotypic leap’, *BioEssays* **34**(10), 893–900.
- Covo, S., Puccia, C. M., Argueso, J. L., Gordenin, D. A. and Resnick, M. A. (2014), ‘The sister chromatid cohesion pathway suppresses multiple chromosome gain and chromosome amplification.’, *Genetics* **196**(2), 373–384.
- Cranmer, K., Brehmer, J. and Louppe, G. (2020), ‘The frontier of simulation-based inference’, *Proc. Natl. Acad. Sci. U. S. A.* p. 201912789.
- Dhar, R., Sägesser, R., Weikert, C., Yuan, J. and Wagner, A. (2011), ‘Adaptation of *Saccharomyces cerevisiae* to saline stress through laboratory evolution.’, *J. Evol. Biol.* **24**(5), 1135–53.

- Dunham, M. J., Badrane, H., Ferea, T., Adams, J., Brown, P. O., Rosenzweig, F. and Botstein, D. (2002), ‘Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*’, *Proc. Natl. Acad. Sci.* **99**(25), 16144–16149.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013), *Bayesian Data Analysis, Third Edition*, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis.
- Gerstein, A. C. and Berman, J. (2018), ‘Diversity of acquired adaptation to fluconazole is influenced by genetic background and ancestral fitness in *Candida albicans*’, *bioRxiv* p. 360347.
- Gerstein, A. C., Ono, J., Lo, D. S., Campbell, M. L., Kuzmin, A. and Otto, S. P. (2015), ‘Too much of a good thing: the unique and repeated paths toward copper adaptation.’, *Genetics* **199**(2), 555–71.
- Gresham, D., Desai, M. M., Tucker, C. M., Jenq, H. T., Pai, D. A., Ward, A., DeSevo, C. G., Botstein, D. and Dunham, M. J. (2008), ‘The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast’, *PLoS Genet.* **4**(12).
- Hartfield, M. and Keightley, P. D. (2012), ‘Current hypotheses for the evolution of sex and recombination.’, *Integr. Zool.* **7**(2), 192–209.
- Hong, J. and Gresham, D. (2014), ‘Molecular Specificity, Convergence and Constraint Shape Adaptive Evolution in Nutrient-Poor Environments’, *PLoS Genet.* **10**(1).
- Kass, R. E. and Raftery, A. E. (1995), ‘Bayes Factors’, *J. Am. Stat. Assoc.* **90**(430), 773.
- Kasuga, T., Bui, M., Bernhardt, E., Swiecki, T., Aram, K., Cano, L. M., Webber, J., Brasier, C., Press, C., Grünwald, N. J., Rizzo, D. M. and Garbelotto, M. (2016), ‘Host-induced aneuploidy and phenotypic diversification in the Sudden Oak Death pathogen *Phytophthora ramorum*’, *BMC Genomics* **17**(1), 1–17.
- Klinger, E. and Hasenauer, J. (2017), A Scheme for Adaptive Selection of Population Sizes in Approximate Bayesian Computation - Sequential Monte Carlo, in J. Feret and H. Koepll, eds, ‘Computational Methods in Systems Biology’, Vol. 10545, Springer International Publishing, pp. 128–144. Series Title: Lecture Notes in Computer Science.
- Klinger, E., Rickert, D. and Hasenauer, J. (2018), ‘pyABC: distributed, likelihood-free inference’, *Bioinformatics* (May), 1–3.
- Lynch, M. (2010), ‘Evolution of the mutation rate.’, *Trends Genet.* **26**(8), 345–352.
- Mannaert, A., Downing, T., Imamura, H. and Dujardin, J. C. (2012), ‘Adaptive mechanisms in pathogens: Universal aneuploidy in *Leishmania*’, *Trends Parasitol.* **28**(9), 370–376.

- Möller, M., Habig, M., Freitag, M. and Stukenbrock, E. H. (2018), ‘Extraordinary Genome Instability and Widespread Chromosome Rearrangements During Vegetative Growth’, *Genetics* **210**(2), 517–529.
- Musacchio, A. and Salmon, E. D. (2007), ‘The spindle-assembly checkpoint in space and time’, *Nat. Rev. Mol. Cell Biol.* **8**(5), 379–393.
- Naylor, R. M. and van Deursen, J. M. (2016), ‘Aneuploidy in Cancer and Aging’, *Annu. Rev. Genet.* **50**(1), 45–66.
- Nei, M. (2013), *Mutation-Driven Evolution*, 1st edn, Oxford University Press, Oxford.
- Niwa, O., Tange, Y. and Kurabayashi, A. (2006), ‘Growth arrest and chromosome instability in aneuploid yeast’, *Yeast* **23**(13), 937–950.
- Otto, S. P. (2009), ‘The Evolutionary Enigma of Sex’, *Am. Nat.* **174**(July), S1–S14.
- Otto, S. P. and Day, T. (2007), *A biologist’s guide to mathematical modeling in ecology and evolution*, Princeton University Press.
- Pavelka, N., Rancati, G., Zhu, J., Bradford, W. D., Saraf, A., Florens, L., Sanderson, B. W., Hattem, G. L. and Li, R. (2010), ‘Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast.’, *Nature* **468**(7321), 321–5.
- Ram, Y., Dellus-Gur, E., Bibi, M., Karkare, K., Obolski, U., Feldman, M. W., Cooper, T. F., Berman, J. and Hadany, L. (2019), ‘Predicting microbial growth in a mixed culture from growth curve data’, *Proc. Natl. Acad. Sci. U. S. A.* **116**(29), 14698–14707.
- Rancati, G. and Pavelka, N. (2013), ‘Karyotypic changes as drivers and catalyzers of cellular evolvability: A perspective from non-pathogenic yeasts’, *Semin. Cell Dev. Biol.* **24**(4), 332–338.
- Rancati, G., Pavelka, N., Fleharty, B., Noll, A., Trimble, R., Walton, K., Perera, A., Staehling-Hampton, K., Seidel, C. W. and Li, R. (2008), ‘Aneuploidy Underlies Rapid Adaptive Evolution of Yeast Cells Deprived of a Conserved Cytokinesis Motor’, *Cell* **135**(5), 879–893.
- Robbins, N., Caplan, T. and Cowen, L. E. (2017), ‘Molecular Evolution of Antifungal Drug Resistance’, *Annu. Rev. Microbiol.* **71**(1), 753–775.
- Rodrigues, M. L. and Albuquerque, P. C. (2018), ‘Searching for a change: The need for increased support for public health and research on fungal diseases’, *PLoS Negl. Trop. Dis.* **12**(6), 1–5.

- Sansregret, L. and Swanton, C. (2017), ‘The Role of Aneuploidy in Cancer Evolution’, *Cold Spring Harb. Perspect. Med.* **7**(1), a028373.
- Santaguida, S. and Amon, A. (2015), ‘Short- and long-term effects of chromosome mis-segregation and aneuploidy’, *Nat. Rev. Mol. Cell Biol.* **16**(8), 473–485.
- Santaguida, S., Vasile, E., White, E. and Amon, A. (2015), ‘Aneuploidy-induced cellular stresses limit autophagic degradation’, *Genes Dev.* **29**(19), 2010–2021.
- Schvartzman, J. M., Sotillo, R. and Benezra, R. (2010), ‘Mitotic chromosomal instability and cancer: Mouse modelling of the human disease’, *Nat. Rev. Cancer* **10**(2), 102–115.
- Selmecki, A. M., Dulmage, K., Cowen, L. E., Anderson, J. B. and Berman, J. (2009), ‘Acquisition of Aneuploidy Provides Increased Fitness during the Evolution of Antifungal Drug Resistance’, *PLoS Genet.* **5**(10), e1000705.
- Selmecki, A. M., Forche, A. and Berman, J. (2010), ‘Genomic Plasticity of the Human Fungal Pathogen *Candida albicans*’, *Eukaryot. Cell* **9**(7), 991–1008.
- Selmecki, A. M., Gerami-Nejad, M., Paulson, C., Forche, A. and Berman, J. (2008), ‘An isochromosome confers drug resistance in vivo by amplification of two genes, ERG11 and TAC1’, *Mol. Microbiol.* **68**(3), 624–641.
- Sheltzer, J. M. and Amon, A. (2011), ‘The aneuploidy paradox: Costs and benefits of an incorrect karyotype’, *Trends Genet.* **27**(11), 446–453.
- Sheltzer, J. M., Ko, J. H., Replogle, J. M., Habibe Burgos, N. C., Chung, E. S., Meehl, C. M., Sayles, N. M., Passerini, V., Storchova, Z. and Amon, A. (2017), ‘Single-chromosome Gains Commonly Function as Tumor Suppressors’, *Cancer Cell* **31**(2), 240–255.
- Shor, E. and Perlin, D. S. (2015), ‘Coping with Stress and the Emergence of Multidrug Resistance in Fungi’, *PLOS Pathog.* **11**(3), e1004668.
- Sionov, E., Lee, H., Chang, Y. C. and Kwon-Chung, K. J. (2010), ‘*Cryptococcus neoformans* Overcomes Stress of Azole Drugs by Formation of Disomy in Specific Multiple Chromosomes’, *PLoS Pathog.* **6**(4), e1000848.
- Sisson, S. A., Fan, Y. and Tanaka, M. M. (2007), ‘Sequential Monte Carlo without likelihoods’, *Proc. Natl. Acad. Sci.* **104**(6), 1760–1765.
- Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M. and Dessimoz, C. (2013), ‘Approximate Bayesian Computation’, *PLoS Comput. Biol.* **9**(1), e1002803.

Sunshine, A. B., Payen, C., Ong, G. T., Liachko, I., Tan, K. M. and Dunham, M. J. (2015), ‘The Fitness Consequences of Aneuploidy Are Driven by Condition-Dependent Gene Effects’, *PLOS Biol.* **13**(5), e1002155.

Todd, R. T., Forche, A. and Selmecki, A. M. (2017), ‘Ploidy Variation in Fungi: Polyploidy, Aneuploidy, and Genome Evolution’, *Microbiol. Spectr.* **5**(4), 1–20.

Torres, E. M., Dephoure, N., Panneerselvam, A., Tucker, C. M., Whittaker, C. A., Gygi, S. P., Dunham, M. J. and Amon, A. (2010), ‘Identification of aneuploidy-tolerating mutations’, *Cell* **143**(1), 71–83.

Torres, E. M., Sokolsky, T., Tucker, C. M., Chan, L. Y., Boselli, M., Dunham, M. J. and Amon, A. (2007), ‘Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast’, *Science* (80-). **317**(5840), 916–924.

Tsai, H. J., Nelliat, A. R., Choudhury, M. I., Kucharavy, A., Bradford, W. D., Cook, M. E., Kim, J., Mair, D. B., Sun, S. X., Schatz, M. C. and Li, R. (2019), ‘Hypo-osmotic-like stress underlies general cellular defects of aneuploidy’, *Nature* .

Williams, B. R., Prabhu, V. R., Hunter, K. E., Glazier, C. M., Whittaker, C. a., Housman, D. E. and Amon, A. (2008), ‘Aneuploidy Affects Proliferation and Spontaneous Immortalization in Mammalian Cells’, *Science* (80-). **322**(5902), 703–709.

Yona, A., Frumkin, I. and Pilpel, Y. (2015), ‘A Relay Race on the Evolutionary Adaptation Spectrum’, *Cell* **163**(3), 549–559.

Yona, A., Manor, Y. S., Herbst, R. H., Romano, G. H., Mitchell, A., Kupiec, M., Pilpel, Y. and Dahan, O. (2012), ‘Chromosomal duplication is a transient evolutionary solution to stress.’, *Proc. Natl. Acad. Sci.* **109**(51), 21010–5.

Zhu, J., Pavelka, N., Bradford, W. D., Rancati, G. and Li, R. (2012), ‘Karyotypic determinants of chromosome instability in aneuploid budding yeast’, *PLoS Genet.* **8**(5).

Zhu, J., Tsai, H.-J., Gordon, M. R. and Li, R. (2018), ‘Cellular Stress Associated with Aneuploidy’, *Dev. Cell* **44**(4), 420–431.

Zhu, Y. O., Sherlock, G. and Petrov, D. A. (2016), ‘Whole Genome Analysis of 132 Clinical *Saccharomyces cerevisiae* Strains Reveals Extensive Ploidy Variation’, *G3 Genes, Genomes, Genet.* **6**(8), 2421–2434.

Zhu, Y. O., Siegal, M. L., Hall, D. W. and Petrov, D. A. (2014), ‘Precise estimates of mutation rate

and spectrum in yeast', **111**(22), E2310–E2318.

URL: <http://www.pnas.org/cgi/doi/10.1073/pnas.1323011111>

Supplementary Material

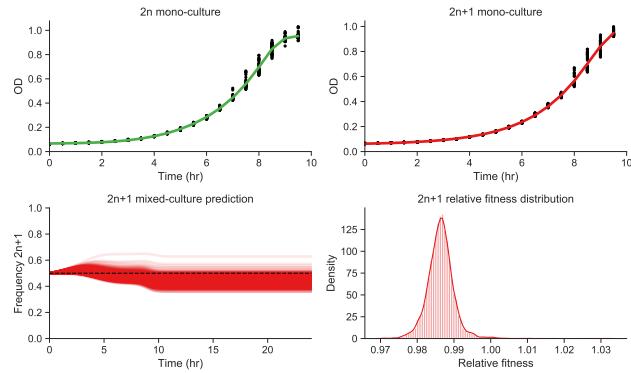


Figure S1: Fitness estimation from 30 °C.

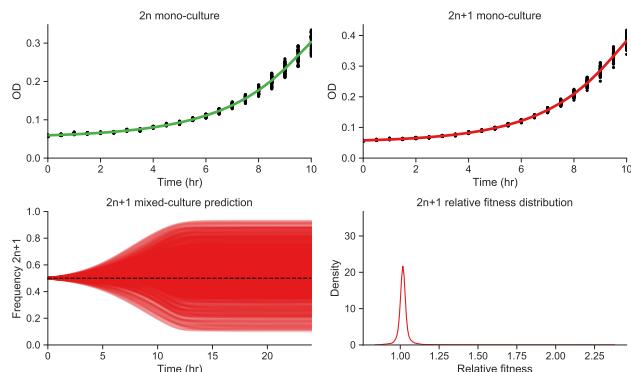


Figure S2: Fitness estimation from 39 °C.

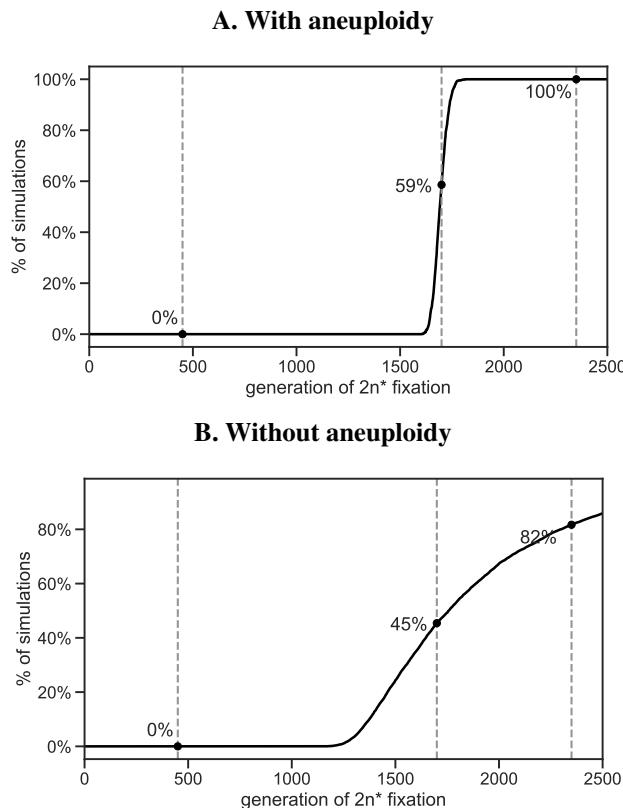


Figure S3: Single-locus model fit. Cumulative distribution of 10000 simulations are shown. **(A)** Simulations parameters are from the maximum a posterior estimation of the single-locus model fit. The likelihood for the heat-stress experiment (Equation 4) is 0.86. **(B)** Simulations parameters are from the maximum a posterior estimation of the single-locus model *without* aneuploidy fit. The likelihood for the heat-stress experiment (Equation 6) is 0.75.

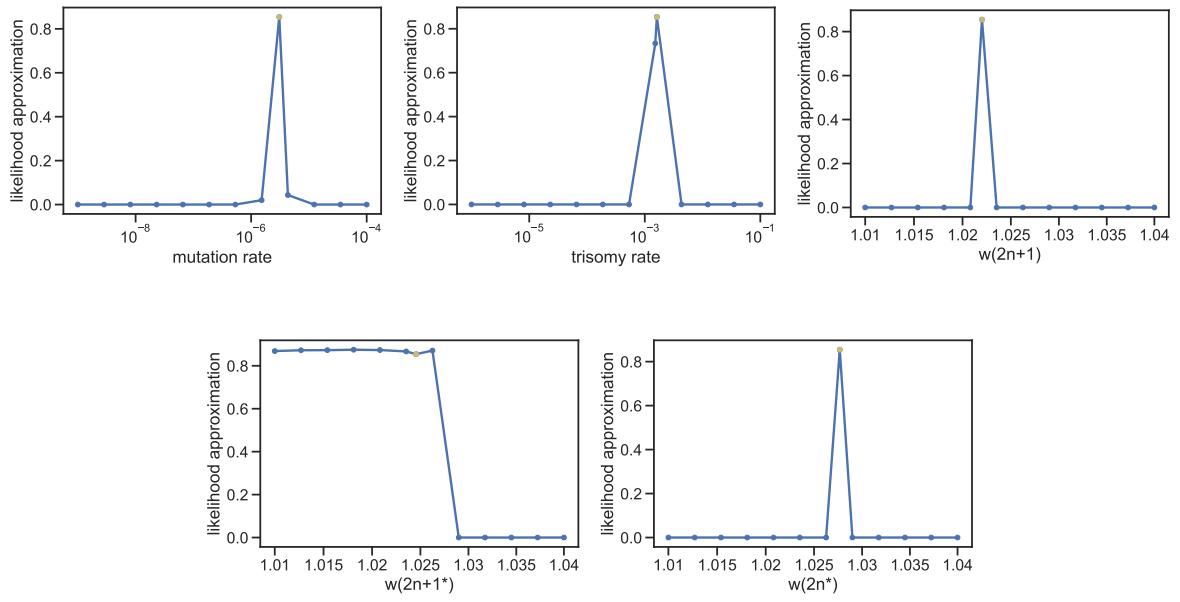


Figure S4: Sensitivity analysis. We take parameters that are maximum a posterior (MAP) for single locus model. Then we change only one parameter at time while measuring the model likelihood of the parameters. x-axis label represent the parameter that we change. Yellow point indicates the MAP value.

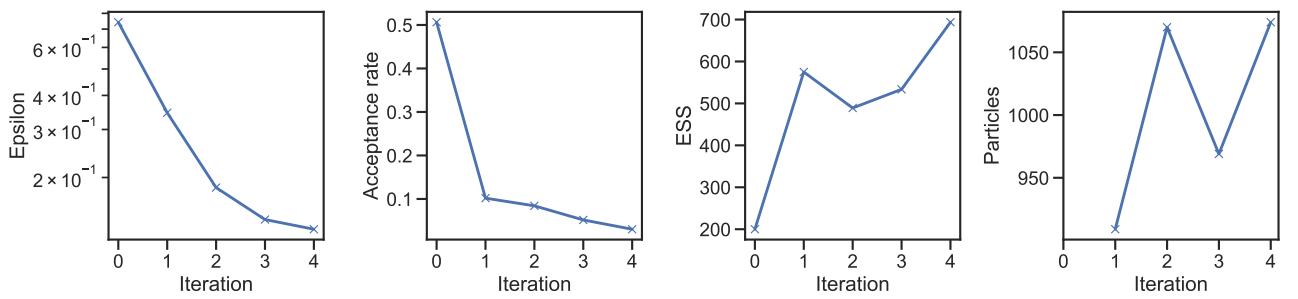


Figure S5: Single-locus model convergence. Pyabc run over four iterations. The model is converged when acceptance threshold (epsilon) is 0.13, the number of particles (accepted instances) is 1074 giving the effective sample size (ESS) 700.

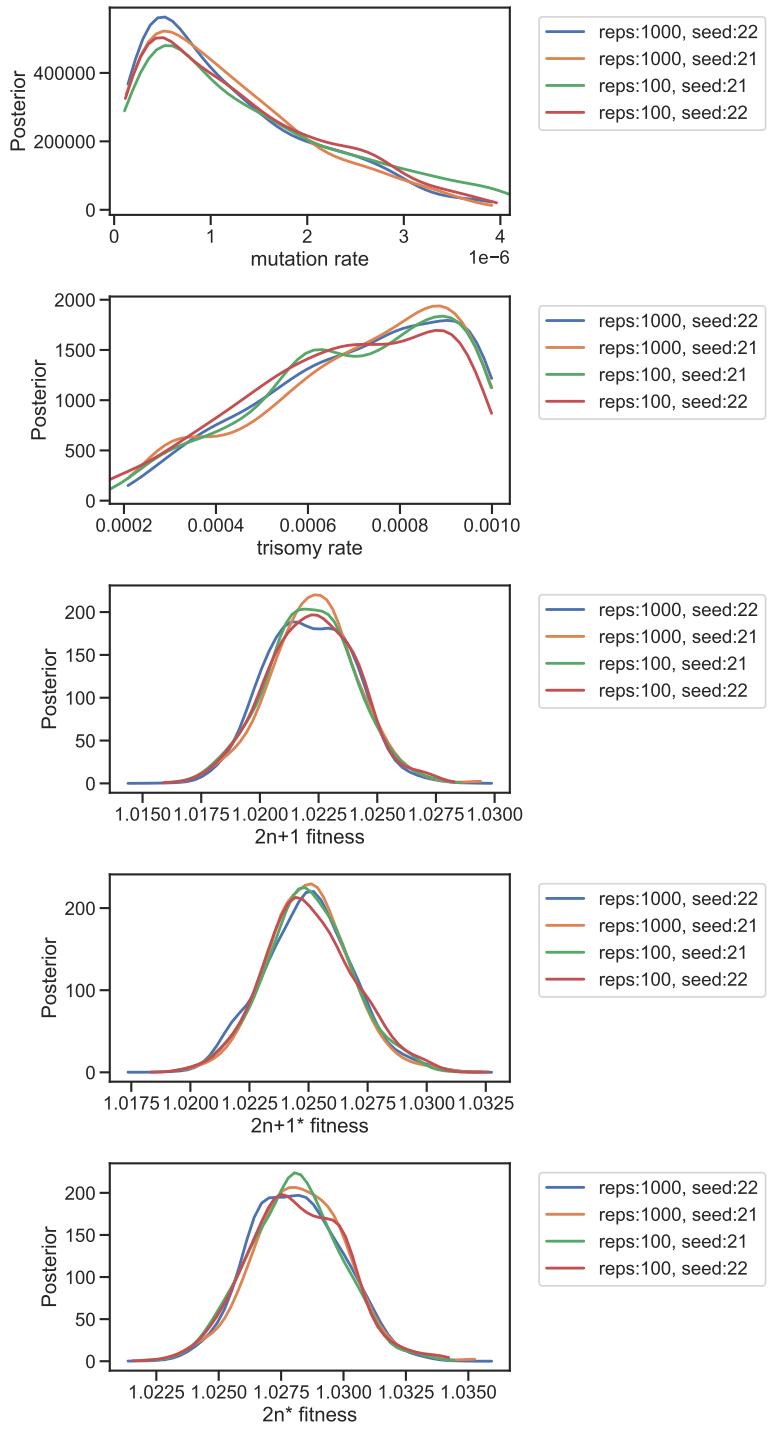


Figure S6: Posterior for different runs For each parameter of the single-locus model, the posterior is represented for runs with different initialization seeds and different number of replicas in simulations run.