

# Adaptive evolution with aneuploidy and mutation

Ilia Kohanovski<sup>1,2,\*</sup>, Martin Pontz<sup>1,\*</sup>, Avihu H. Yona<sup>3</sup>, and Yoav Ram<sup>1,†</sup>

<sup>1</sup>School of Zoology, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

<sup>2</sup>School of Computer Science, Reichman University, Herzliya, Israel

<sup>3</sup>Institute 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

June 22, 2022

## 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. Importantly, aneuploidy differs from point mutations in rate, fitness effect, and reversibility. Here, we develop an evolutionary model for adaptive evolution with both aneuploidy and mutation, and apply it within an Bayesian inference framework. We estimate the formation rate and fitness effect of aneuploidy and mutation in an evolutionary experiment in which *Saccharomyces cerevisiae* adapted to heat stress. Our results suggest that aneuploidy is more frequent than previously estimated and that experimental observations cannot be explained without aneuploidy. We do not find evidence for genetic instability in aneuploid cells. Our results support the hypothesis that aneuploidy is a transient adaptive hypothesis: the experimental populations adapted by first acquiring a chromosome duplication; then by acquiring a beneficial mutation that likely appeared in aneuploid cells, rather than euploid cells; and finally by rapidly losing the extra chromosome to avoid paying the fitness cost of aneuploidy.

# 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 limi-

tation (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 an evolutionary-genetic model that includes the effects of natural selection, genetic drift, aneuploidy, and mutation to examine the role of aneuploidy in adaptive evolution. This model follows a population of cells characterised by both their ploidy and their genotype. We fit this model to the experimental results of Yona et al. (2012) using an *approximate Bayesian computation* framework (Sisson et al., 2007) to infer model parameters, including rates and fitness effects of aneuploidy and mutation, and to test hypotheses about the evolutionary process by comparing different versions of the model and predicting the genotype frequency dynamics using the estimated parameter values.

## Models and Methods

**Evolutionary Model.** We model the evolution of a population of cells using a Wright-Fisher model (Otto and Day, 2007), assuming a constant effective population size  $N$ , non-overlapping

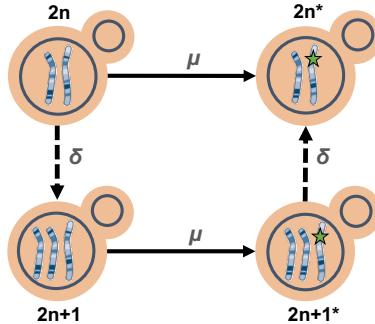
generations, and including the effects of natural selection, genetic drift, aneuploidy, and mutation. We focus on beneficial genetic modifications, neglecting the effects of deleterious and neutral mutations or karyotypic changes. The model allows for a single aneuploid karyotype (e.g., chromosome III duplication) and a single mutation to accumulate in the genotype. Thus, the model follows four genotypes (Figure 1): euploid wild-type,  $2n$ , the initial genotype; euploid mutant,  $2n^*$ , with the standard karyotype and a single beneficial mutation; aneuploid wild-type,  $2n+1$ , with an extra chromosome, i.e., following chromosome duplication; and aneuploid mutant,  $2n+1^*$ , with an extra chromosome and a beneficial mutation.

Transitions between the genotypes occur as follows (Figure 1): Beneficial mutations from  $2n$  to  $2n^*$  and from  $2n+1$  to  $2n+1^*$  occur with probability  $\mu$ , the mutation rate. We neglect back-mutations (i.e., from  $2n^*$  to  $2n$  and from  $2n+1^*$  to  $2n+1$ ). Aneuploidy is formed by chromosome mis-segregation, so that cells transition from  $2n$  to  $2n+1$  and from  $2n+1^*$  to  $2n^*$  with probability  $\delta$ , the aneuploidy rate. That is, we assume chromosomes are gained and lost at the same rate, and we neglect events that form a less-fit genotype (i.e.,  $2n+1$  to  $2n$  and  $2n^*$  to  $2n+1^*$ ).

In the experiment by Yona et al. (2012), the population was grown every day from  $1.6 \cdot 10^6$  cells until reaching stationary phase and then diluted 1:120. Thus, we set the population size to  $N = 6.425 \cdot 10^6$ , the harmonic mean of  $\{2^k \cdot 1.6 \cdot 10^6\}_{k=0}^7$  (Crow and Kimura, 1970). The initial population has  $N$  cells with genotype  $2n$ . The effect of natural selection on the frequency  $f_i$  of genotype  $i = 2n, 2n+1, 2n+1^*, \text{ or } 2n^*$  is given by

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

where  $w_i$  is the fitness of genotype  $i$  and  $\bar{w} = \sum_j f_j w_j$  is the population mean fitness. The effect of



**Figure 1: Model Illustration.** There are four genotypes in our model: euploid wild-type,  $2n$ ; euploid mutant,  $2n^*$ ; aneuploid wild-type,  $2n+1$ ; and aneuploid mutant,  $2n+1^*$ . Overall there are two possible trajectories from  $2n$  to  $2n^*$ . Arrows denote transitions between genotypes, with transition rates  $\mu$  for the beneficial mutation rate and  $\delta$  for the aneuploidy rate.

mutation and aneuploidy on genotype frequencies is given by

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

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

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

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  $\text{Mult}(N, \mathbf{f})$  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  $f_i \rightarrow f'_i$ .

**Empirical evidence.** We use the results of evolutionary experiments reported by Yona et al. (2012). In their 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). From re-analysis of data not published in the original paper, aneuploidy did not fix before at least 200 generations elapsed. The experiment continued with two populations, in which aneuploidy was eliminated by generation 1,700 and 2,350 while still under the same conditions of elevated heat (39 °C).

**Likelihood function.** Because our model, just like the Wright-Fisher model, is non-linear and stochastic, computing the distribution of fixation time  $T(g)$  of genotype  $g$  for use in the likelihood function is intractable (it is even hard to use a diffusion-equation approximation due to the model having multiple genotypes, rather than just two). We overcome this problem by approximating the likelihood using simulations. We simulate 1,000 experiments per parameter vector  $\theta = (\mu, \delta, s, b, c)$ , resulting in a set of simulated observations  $\tilde{\mathbf{X}} = \{\tilde{X}_i\}_{i=1}^{1000}$ . We then compute the approximate likelihood,

$$\begin{aligned} \mathcal{L}(\theta) &= P^4(200 \leq T(2n+1) \leq 450) \cdot \left[ 1 - \right. \\ &\quad P_{\tilde{\mathbf{X}}}^4(\{T(2n^*) < 1700\} \mid 200 \leq T(2n+1) \leq 450) - \\ &\quad P_{\tilde{\mathbf{X}}}^4(\{1700 < T(2n^*) < 2350\} \mid 200 \leq T(2n+1) \leq 450) + \\ &\quad \left. P_{\tilde{\mathbf{X}}}^4(\{T(2n^*) < 1700\} \wedge \{1700 < T(2n^*) < 2350\} \mid 200 \leq T(2n+1) \leq 450) \right], \end{aligned} \tag{4}$$

where  $!\{\dots\}$  is the "logical not" operator,  $P^4(\dots)$  is the 4th power of  $P(\dots)$ , and all probabilities  $P_{\tilde{\mathbf{X}}}(\dots)$  are approximated from the results of the simulations  $\tilde{\mathbf{X}}$ . For example,  $P_{\tilde{\mathbf{X}}}(!\{T(2n^*) < 1700\} \mid 200 \leq T(2n+1) \leq 450)$  is approximated by taking simulations in which  $2n+1$  fixed before generation 450 but not before generation 200, and computing the fraction of such simulations in which  $2n^*$  did not fix by generation 1,700, and hence aneuploidy did not extinct before generation 1,700. Figure S1 compares results with less and more simulated experiments, demonstrating that 1,000 simulations are likely sufficient.

For a model without aneuploidy (that is, when the aneuploidy rate is fixed at zero,  $\delta = 0$ ), we disregard the increased expression in chromosome III and the growth advantage measured in generation 450, and focus on the growth advantage measured in later generations, presumably due to a beneficial mutation. Therefore, the likelihood is approximated by

$$\begin{aligned}\mathcal{L}_!(\theta) = 1 - P_{\tilde{\mathbf{X}}}^4(!\{T(2n^*) < 1700\}) - \\ P_{\tilde{\mathbf{X}}}^4(!\{1700 < T(2n^*) < 2350\}) + \\ P_{\tilde{\mathbf{X}}}^4(!\{T(2n^*) < 1700\} \wedge !\{1700 < T(2n^*) < 2350\}) .\end{aligned}\quad (5)$$

**Parameter inference.** To infer model parameters, we use approximate Bayesian computation with a sequential Monte-Carlo scheme, or ABC-SMC (Sisson et al., 2007), implemented in the pyABC Python package (Klinger et al., 2018, [pyabc.readthedocs.io](#)). 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 directly computed. It is therefore suitable in our case, in which the likelihood function can only be approximated from simulations, and cannot be directly computed.

The ABC-SMC algorithm employs sequential importance sampling over multiple iterations (Toni et al., 2009; Klinger and Hasenauer, 2017; Syga et al., 2021). In iteration  $t$  of the algorithm, a set of parameter vectors,  $\{\theta_{i,t}\}_{i=1}^{n_t}$ , also called *particles*, are constructed in the following way. A proposal particle,  $\theta^*$ , is sampled from a proposal distribution, and is either accepted or rejected, until  $n_t$  particles are accepted. The number of particles,  $n_t$ , is adapted at every iteration  $t$  using the adaptive population strategy (Klinger et al., 2018, [pyabc.readthedocs.io](#)). For  $t = 0$ , the proposal particle is sampled from the prior distribution,  $p(\theta)$ . For  $t > 0$ , the proposal particle is sampled from the particles accepted in the previous iteration,  $\{\theta_{i,t-1}\}_{i=1}^{n_{t-1}}$ , each with a probability relative to its weight  $W_{t-1}(\theta_{i,t-1})$  (see below). The proposal particle is then perturbed using a kernel perturbation kernel,  $K_t(\theta^* \mid \theta)$  where  $\theta$  is the sample from the previous iteration. Then, a set of synthetic observations

$\tilde{\mathbf{X}}^*$  is simulated, and the proposal particle  $\theta^*$  is accepted if its approximate likelihood (eq. (4)) is high enough,  $\mathcal{L}(\theta^*) > 1 - \epsilon_t$  (or more commonly, if  $1 - \mathcal{L}(\theta^*) < \epsilon_t$ ), where  $\epsilon_t > 0$  is the *acceptance threshold*, as higher values of  $\epsilon_t$  allow more particles to be accepted. The acceptance threshold  $\epsilon_t$  is chosen as the median of the  $1 - \mathcal{L}(\theta)$  of the particles accepted in the previous iteration,  $t - 1$ , and  $\epsilon_0 = 0.01$ . For each accepted particle  $\theta_{i,t}$  a weight  $W_t(\theta_{i,t})$  is assigned: for  $t = 0$ ,  $W_0(\theta_{i,0}) = 1$ , and for  $t > 0$ ,  $W_t(\theta_{i,t}) = p(\theta_{i,t}) / \sum_{i=1}^{n_{t-1}} W_{t-1}(\theta_{i,t-1}) K_t(\theta_{i,t}, \theta_{i,t-1})$ , where  $p(\theta)$  is the prior density of  $\theta$  and  $K_t(\theta', \theta)$  is the probability of a perturbation from  $\theta$  to  $\theta'$ .  $K_t(\theta' | \theta)$  is a multivariate normal distribution, fitted at iteration  $t$  to the particles from the previous iteration,  $\{\theta_{i,t-1}\}_{i=1}^{n_{t-1}}$ , and their weights,  $\{W(\theta_{i,t-1})\}_{i=1}^{n_{t-1}}$ .

Acceptance is determined according to the approximate likelihood (eq. (4)), which has a maximum value of  $\mathcal{L}_{max} = 0.875$  (giving a minimal value of  $\epsilon_{min} = 0.125$ ). We terminated the inference iterations when the change in  $\epsilon$  value from one iteration to the next was small. With our standard prior and model, we reached  $\epsilon = 0.13$  (or  $\mathcal{L} = 0.87$ ) after six iterations, with  $n_6 = 982$  accepted parameter vectors and effective sample size ESS=651 (Figure S2). Running the inference algorithm with different initialization seeds and less or more simulations for approximating the likelihood produced similar posterior distributions (Figure S1).

After producing a set of weighted particles from the the posterior distribution using the above ABC-SMC algorithm, we approximate the posterior using kernel density estimation (KDE) with Gaussian kernels. Since KDE can have positive posterior probability for values with the prior zero, we truncate KDE, giving zero probability for the ranges outside of the prior. From maximum of the KDE function we get the MAP (maximum a posteriori) estimate, and then draw 5,000,000 samples from the posterior KDE to compute the HDI (highest density interval) and draw 50,000 samples from it to visualize the posterior distribution with histograms.

**Model comparison.** We examine several versions of our evolutionary models, e.g. without aneuploidy or with increased mutation rate in aneuploid cells, as well as several different prior distributions (see below). To compare these, we plot posterior predictions: for each model we execute 10,000 simulations using the MAP parameter estimates and plot the distributions of time to fixation of  $2n^*$ , one of key properties of the model likelihood. These plots visualize the fit of each model to the data. Also, for similar models we plot the marginal and joint posterior distributions of the parameters; if these are similar, we consider the models interchangeable. We validate this by comparing HDI (highest density interval) of posterior distributions.

Where posterior plots are very similar and the number of parameters is the same, we use WAIC, or the widely applicable information criterion (Gelman et al., 2013), defined as

$$WAIC(\theta) = -2 \log \mathbb{E}[\mathcal{L}(\theta)] + 2\mathbb{V}[\log \mathcal{L}(\theta)] \quad (6)$$

where  $\theta$  is a parameter vector, and  $\mathbb{E}[\cdot]$  and  $\mathbb{V}[\cdot]$  are the expectation and variance taken over the posterior distribution, which in practice are approximated using 50,000 samples from the posterior KDE. We validated that upon resampling WAIC values do not significantly change and that differences in WAIC between models are preserved. WAIC values are scaled as a deviance measure: lower values imply higher predictive accuracy (Kass and Raftery, 1995).

**Prior distributions.** We used informative prior distributions for  $w_{2n+1} = 1 - c + b$ ,  $w_{2n+1^*} = (1 + s)(1 - c) + b$  and  $w_{2n^*} = 1 + s$ , which we estimated from growth curves data from mono-culture growth experiments previously reported by Yona et al. (2012, Figs. 3C, 4A, and S2). We used Curveball, a method for predicting results of competition experiments from growth curve data (Ram et al., 2019, [curveball.yoavram.com](http://curveball.yoavram.com)). Briefly, 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. Because Curveball uses a maximum-likelihood approach to estimate model parameters, we were able to estimate a distribution of relative fitness values to be used as a prior distribution by sampling 10,000 samples from a truncated multivariate normal distribution defined by the maximum-likelihood covariance matrix (Figure S3).

We used growth curves of  $2n$  and  $2n+1$  in 39 °C to estimate an informative prior distribution for  $w_{2n+1}$  (Figure S3-D, assuming  $w_{2n} = 1$ ). In this prior distribution, we used the same prior for  $w_{2n+1^*}$  and  $w_{2n^*}$ . To increase computational efficiency, we also assumed  $w_{2n^*} > w_{2n+1^*} > w_{2n+1} > w_{2n}$ ; running the inference without this assumption produced similar results. See *supporting material* for an extended informative prior distribution that uses growth curves of  $2n^*$  and  $2n+1$  growing in 39 °C; this prior distribution proved to be less useful.

As a control, we tested an uninformative uniform prior with  $U(1, 6)$ , for (i) all  $w_{2n+1}$ ,  $w_{2n+1^*}$ ,  $w_{2n^*}$ , or (ii) only for  $w_{2n+1^*}$ ,  $w_{2n^*}$ , using the above informative prior for  $w_{2n+1}$ . In these cases the inference algorithm failed to converge.

For the mutation rate,  $\mu$ , and aneuploidy rate,  $\delta$ , we used uninformative uniform priors,  $\mu \sim U(10^{-9}, 10^{-5})$  and  $\delta \sim U(10^{-6}, 10^{-2})$ . A wider mutation rate prior,  $\mu \sim U(10^{-9}, 10^{-3})$ , produced similar results.

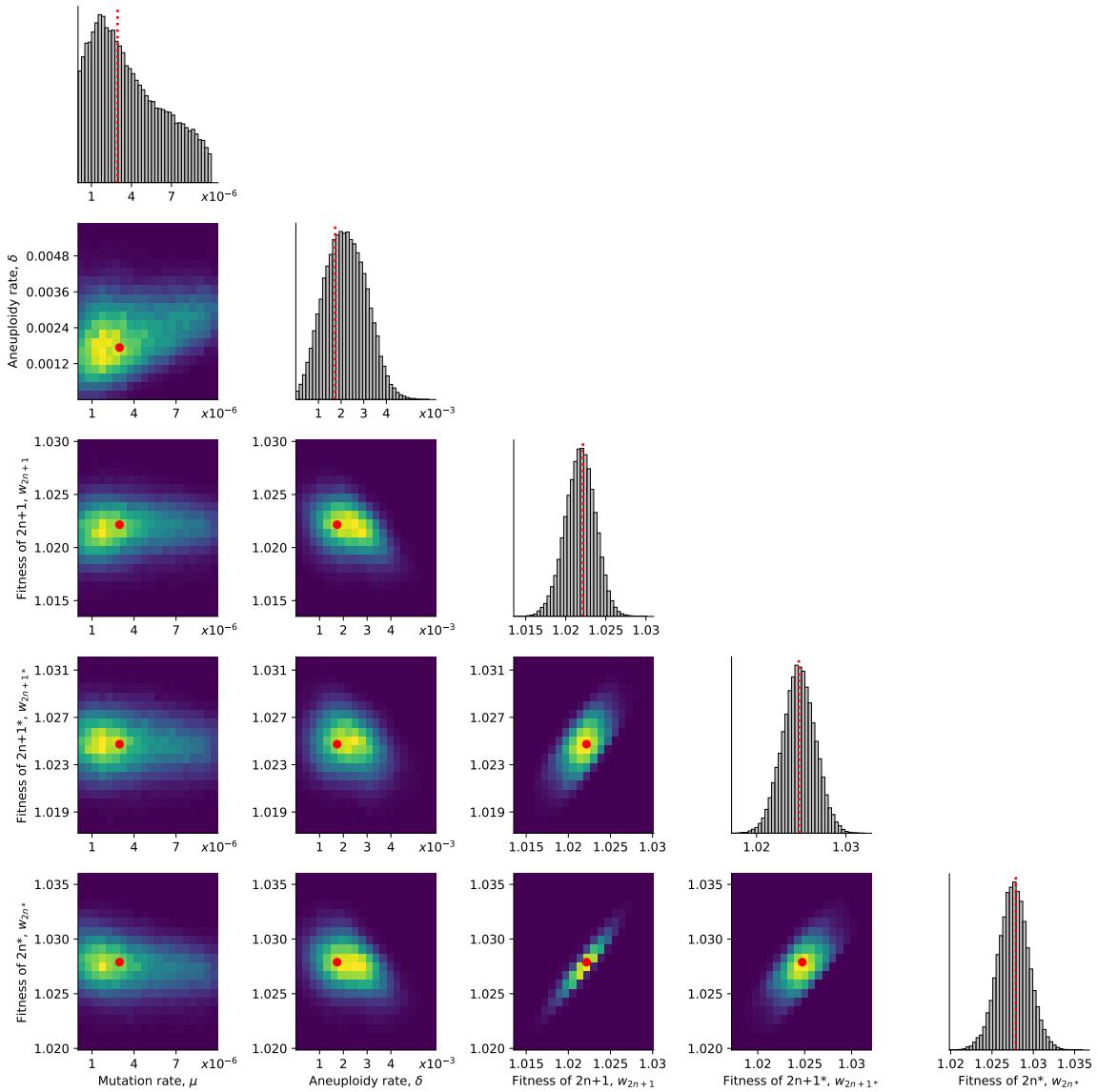
## Results

**Rates and fitness effects of aneuploidy and mutation.** We used ABC-SMC to infer the posterior distribution of model parameters (Figure 2). We report parameter estimates using the MAP (maximum a posteriori) and providing the 50% HDI (highest density interval) in square brackets. See *Supporting material* for sensitivity analysis. The estimated mutation rate,  $\mu = 2.965 \cdot 10^{-6}$  [ $2.718 \cdot 10^{-7} - 3.589 \cdot 10^{-6}$ ], corresponds to a mutation target size of  $\sim 10^4$ , assuming the mutation rate per base pair is roughly  $2 \cdot 10^{-10}$  (Zhu et al., 2014) or  $3.3 \cdot 10^{-10}$  (Lynch et al., 2008). The estimated aneuploidy rate,  $\delta = 1.72 \cdot 10^{-3}$  [ $1.47 \cdot 10^{-3} - 2.786 \cdot 10^{-3}$ ] is much higher than in previous studies: for chromosome III in diploid *S. cerevisiae*, Zhu et al. (2014) estimated  $6.7 \cdot 10^{-6}$  chromosome gain events per generation, and Kumaran et al. (2013) estimate  $3.0 - 4.3 \cdot 10^{-5}$  chromosome loss events per generation (95% confidence interval). The estimated fitness values are  $w_{2n+1} = 1.022$  [ $1.021 - 1.023$ ],  $w_{2n+1*} = 1.025$  [ $1.024 - 1.026$ ],  $w_{2n*} = 1.028$  [ $1.026 - 1.029$ ], all relative to the fitness of  $2n$ , which is set to  $w_{2n} = 1$ . Thus, we can infer that the cost of trisomy is  $c = w_{2n*} - w_{2n+1*} = 0.003$  (or 0.3%) and the benefit of trisomy is  $w_{2n+1} - 1 - c = 0.019$  (1.9%), whereas the benefit of beneficial mutation is  $w_{2n*} - 1 = 0.028$  (2.8%).

**Role of aneuploidy in the evolutionary dynamics.** Our model fits the data well: in simulations using the MAP parameter estimates,  $2n^*$  fixed in 61% of simulations by generation 1,700 and in 100% of simulations by generation 2,350 (Figure 3B).

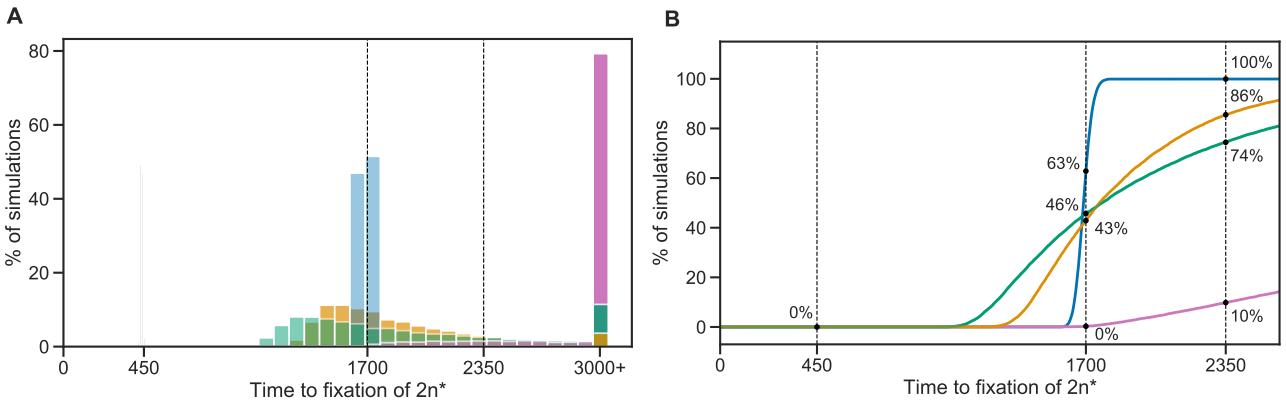
However, a model without aneuploidy (where the aneuploidy rate is fixed at zero,  $\delta = 0$ ), fails to explain the experimental observations (Figure 3). The estimated mutation rate without aneuploidy is  $\mu = 7.98 \cdot 10^{-9}$  [ $7.906 \cdot 10^{-9} - 8.138 \cdot 10^{-9}$ ], much lower compared to a model with aneuploidy and suggesting a target size of just 40. The fitness of the mutant is also much lower at  $w_{2n*} = 1.013$  [ $1.012 - 1.013$ ]. This is because, without aneuploidy, a high mutation rate or fitness effect will lead to faster appearance and fixation of  $2n^*$  than in the experimental observations. Even with these lower estimates, the model fit is worse than that of a model with aneuploidy (Figure 3).

We also checked a model in which aneuploidy occurs but is adaptively neutral compared to the wild type, that is,  $w_{2n+1} = w_{2n}$  and  $w_{2n+1*} = w_{2n*}$  but  $\delta > 0$ . This model fits the data better than the model with no aneuploidy (in which  $\delta = 0$ ), but worse than a model with positive selection for aneuploidy, in which  $w_{2n} < w_{2n+1} < w_{2n+1*} < w_{2n*}$  (Figure 3). Thus, our results suggest that aneuploidy significantly affects the evolutionary dynamics.



**Figure 2: Posterior distribution of model parameters.** On the diagonal, the inferred posterior distribution of each model parameter. Below the diagonal, the inferred joint posterior distribution of pairs of model parameters (dark purple and bright yellow for low and high density, respectively). Red markers and orange lines for the joint MAP estimate (which may differ from the marginal MAP, as the marginal distribution integrates over all other parameters).

**Genetic instability in aneuploid cells.** It has been suggested that aneuploidy increased genetic instability (Sheltzer et al., 2011). Therefore, we inferred model parameters under the assumption that the mutation rate increases in aneuploid cells by a factor  $\tau = 1, 33/32$  (due to an additional chromosome), 2, 5, 10, or 100 (due to genetic instability). We found that the posterior distribution was similar for  $\tau = 1, 33/32, 2$ , and 5 (Figure S4). With  $\tau = 100$ , the estimated mutation rate was about 7-8-fold lower compared to  $\tau = 1$  ( $\mu = 4.094 \cdot 10^{-7}$  [ $6.252 \cdot 10^{-8} - 6.046 \cdot 10^{-7}$ ]) and the aneuploidy rate was about 2-3-fold lower ( $\delta = 0.744 \cdot 10^{-3}$  [ $0.506 \cdot 10^{-3} - 1.827 \cdot 10^{-3}$ ]). With  $\tau = 10$ , the estimated mutation rate was only slightly lower compared to  $\tau = 1$  ( $\mu = 1.67 \cdot 10^{-6}$  [ $2.836 \cdot 10^{-8} - 2.245 \cdot 10^{-6}$ ]).



**Figure 3: Model fit with and without aneuploidy.** The distribution of time to fixation of  $2n^*$  (i.e., adaptation time) in 10,000 simulations using MAP parameters of the model with beneficial aneuploidy (blue;  $\delta > 0$ ,  $w_{2n} < w_{2n+1} < w_{2n+1}^* < w_{2n}^*$ ) compared to alternative models: a model with the same parameter values but without aneuploidy (gray,  $\delta = 0$ , concentrated at  $t = 450$ ); a model fitted to the data assuming no aneuploidy (green,  $\delta = 0$ ); a model fitted to the data assuming neutral aneuploidy (yellow,  $\delta > 0$ ,  $w_{2n+1} = w_{2n}$ ,  $w_{2n+1}^* = w_{2n}^*$ ); and a model with beneficial aneuploidy and an extended prior distribution (pink). In the experiment by Yona et al. (2012), one population lost aneuploidy by generation 1,700 and another by generation 2,350 (dashed lines) but not before generation 450. Thus, the blue distribution has a better fit compared to the other distributions (the gray distribution has a particularly poor fit). The MAP likelihood (eq. (4)) is 0.84, 0.78, 0.67, and 0.14 for the models represented by blue, yellow, green, and pink distributions, respectively. **(A)** Histogram of the time to fixation of  $2n^*$ . The last bin contains all values equal or greater than 3,000. **(B)** Cumulative distribution of the time to fixation.

WAIC is lowest for  $\tau = 33/32$  and  $\tau = 1$ . Therefore, evidence does not support an increase in mutation rate in aneuploid cells, and moreover, unless the increase is strong ( $\tau \geq 10$ ), it does not seem to affect our inference. We also checked the differences in genotype frequency dynamics for different  $\tau$  values. We observe  $\tau = 100$  could be distinguished if accurate data was available for the waiting time until the frequency of  $2n$  to decrease below 95% (Figure S5A) or for waiting time for the frequency of  $2n+1$  to either reach or go below 95% (Figure S5B).

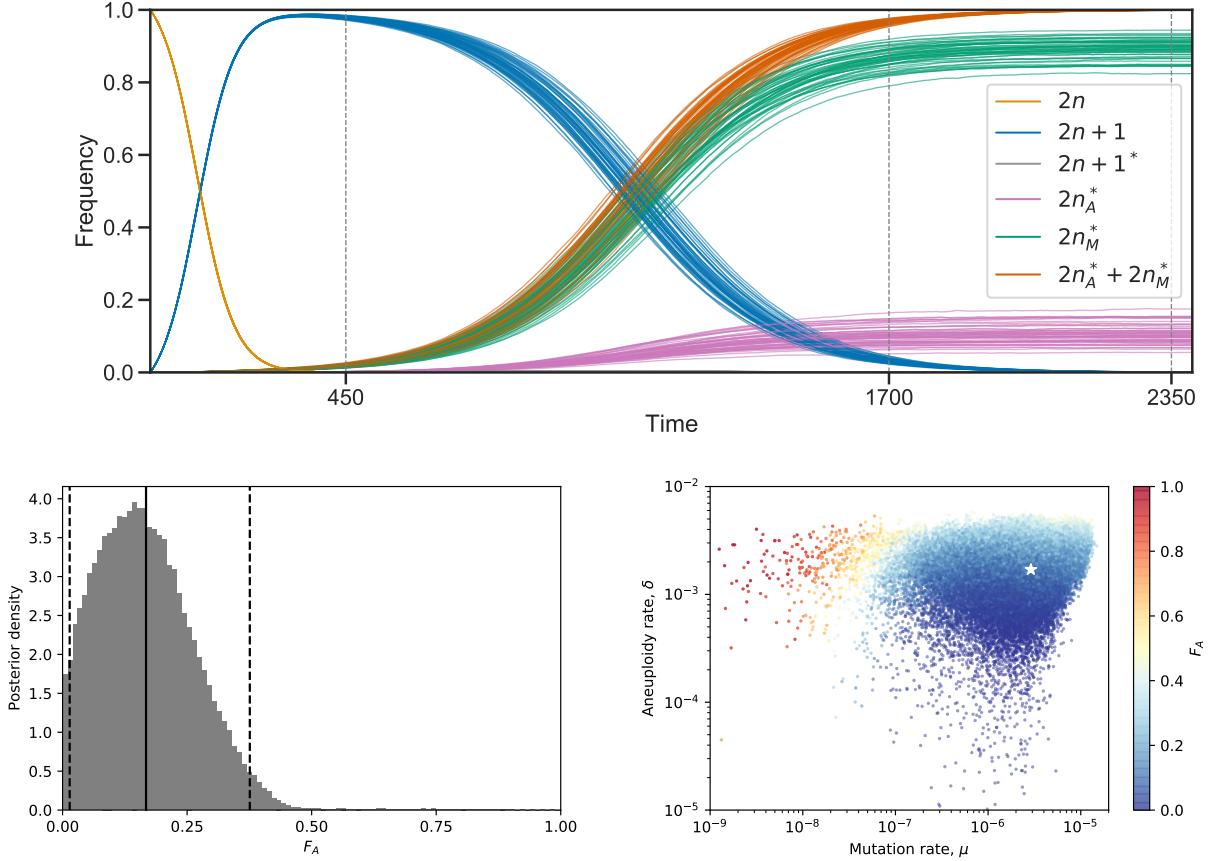
**Competition between mutation and aneuploidy.** We simulated 50 replicate genotype frequency dynamics using the MAP estimate parameters. Figure 4A shows the simulated frequencies of the four genotypes ( $2n$ ,  $2n+1$ ,  $2n+1^*$  and  $2n^*$ ), as well as the frequencies of  $2n^*$  cells that arose from either  $2n+1$  cells via a sequences of mutation and chromosome loss events ( $2n_A^*$ ), or directly from  $2n$  cells via a mutation event ( $2n_M^*$ ). We find that  $2n+1^*$  never reaches substantial frequency as it is quickly replaced by  $2n^*$ .

Yona et al. (2012) have suggested that aneuploidy is a transient adaptive solution, arising quickly due to increased fitness and high formation rate, but ultimately lost due to the cost of aneuploidy as beneficial mutations take over. We tested this hypothesis by estimating  $F_A$ , the expected frequency of  $2n^*$  that arose from  $2n+1$ , as the average frequency of such cells at the end of simulations using the MAP estimate parameters. Surprisingly, we observe that the majority of  $2n^*$  cells are a product of a direct mutation in  $2n$  cells, rather than descending from  $2n+1$  cells ( $F_A^{MAP} = 0.106$ , Figure 4A). This is despite the fact that the  $2n+1$  genotype reaches high frequencies in the population (at least 0.98, Figure 4A).

This result is not unique to the MAP parameter estimate. We simulated genotype frequency dynamics using parameter samples from the posterior distribution (Figure 2), and computed the posterior distribution of  $F_A$  (Figure 4B). The mean  $F_A$  was just 0.1675 [0.0142 - 0.3758 95% CI] and only in 408 of 86,349 posterior samples (0.47%)  $F_A$  was larger than 0.5, that is,  $2n^*$  was more likely to arise from an aneuploid cell. Thus, if we pick up a cell from the evolved  $2n^*$  population, it is more likely to have descended directly from an euploid cell than from an aneuploid cell. The probability of  $2n^*$  descending from  $2n+1$  ( $F_A$ ) increases with the aneuploidy rate,  $\delta$ , and decreases with the mutation rate,  $\mu$ , and in some cases can also be affected by the fitness parameters (Figure 4C and Figure S9).

## Discussion

**Results of current study.** We inferred the rates and fitness effects of aneuploidy and mutation in a previously published evolutionary experiment in which yeast populations adapted to heat stress (Yona et al., 2012). First, we estimate that the aneuploidy rate (i.e., number of chromosome gains per generation) is  $1.7 \cdot 10^{-3}$ , is higher than a previous estimate of  $6.7 \cdot 1e0^{-6}$  (Zhu et al., 2016). Second, the experiments results cannot be explained without aneuploidy ( $\delta = 0$ ) or if aneuploidy is neutral ( $w_{2n+1} = w_{2n}$ ,  $w_{2n+1*} = w_{2n*}$ ), and therefore we conclude that aneuploidy plays a significant role in the dynamics. Third, we find no evidence for increased mutation rates in aneuploid cells, but more data would have allowed a more conclusive statement. Therefore, the effect of aneuploidy on genetic instability remains to be explored in future work. Fourth, we simulated genotype frequency dynamics using inferred rates and fitness effects of aneuploidy and mutation. We found that the aneuploid mutant,  $2n+1^*$ , never reaches high frequency, in a process similar to *stochastic tunneling* (Iwasa et al., 2004). Furthermore, we find that XXXX.



**Figure 4: Posterior genotype frequency dynamics.** (A) Posterior predicted genotype frequencies over time, including the source of  $2n^*$ . Shown are 50 simulations using the MAP estimate parameters. (B) Marginal posterior distribution of  $F_A$ , the expected frequency of  $2n^*$  cells that arose from  $2n+1$  cells, computed as the average frequency at the end of 50 simulations for 86,349 samples from the parameter posterior distribution. Solid and dashed lines show the mean and 95% CI. (C)  $F_A$  values (color coded) from panel B, with their corresponding mutation rate  $\mu$  on x-axis and aneuploidy rate  $\delta$  on the y-axis. White star shows the MAP estimate. See also Figure S9.

**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.
- Crow, J. F. and Kimura, M. (1970), *An introduction to population genetics theory*, Burgess Pub. Co., Minneapolis.

- 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).
- Iwasa, Y., Michor, F. and Nowak, M. A. (2004), ‘Stochastic Tunnels in Evolutionary Dynamics’, *Genetics* **166**(3), 1571–1579.
- 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.
- Kumaran, R., Yang, S.-Y. and Leu, J.-Y. (2013), ‘Characterization of chromosome stability in diploid, polyploid and hybrid yeast cells’, **8**(7), e68094.
- Lynch, M. (2010), ‘Evolution of the mutation rate.’, *Trends Genet.* **26**(8), 345–352.
- Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C. R., Dopman, E. B., Dickinson, W. J., Okamoto, K., Kulkarni, S., Hartl, D. L. and Thomas, W. K. (2008), ‘A genome-wide view of the spectrum of spontaneous mutations in yeast’, **105**(27), 9272–9277.
- 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., Blank, H. M., Pfau, S. J., Tange, Y., George, B. M., Humpton, T. J., Brito, I. L., Hiraoka, Y., Niwa, O. and Amon, A. (2011), ‘Aneuploidy Drives Genomic Instability in Yeast’, *Science (80-. ).* **333**(6045), 1026–1030.

- 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.
- 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.
- Syga, S., David-Rus, D., Schälte, Y., Hatzikirou, H. and Deutsch, A. (2021), ‘Inferring the effect of interventions on COVID-19 transmission networks’, *Sci. Rep.* **11**(1), 1–11.
- 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.
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A. and Stumpf, M. P. (2009), ‘Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems’, *J. R. Soc. Interface* **6**(31), 187–202.
- 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’, *Nature* **111**(22), E2310–E2318.

# Supplementary Material

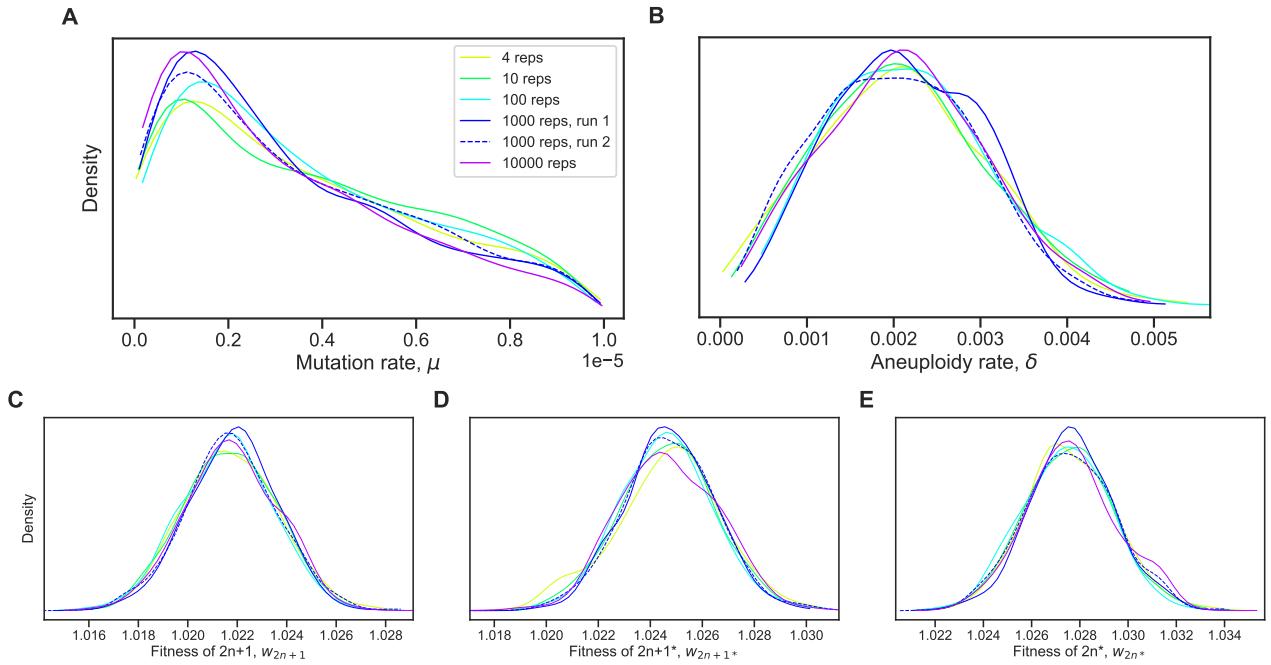
## Supplementary Analysis

**Sensitivity analysis.** Changing a single parameter while keeping the rest fixed at the MAP estimate produces a worse fit to the data (Figure S6). Furthermore, we fitted models with a mutation rate fixed at  $\mu=10^{-5}$ ,  $10^{-6}$  and  $10^{-7}$ . We inferred similar parameters estimates for the model with  $\mu = 10^{-6}$  compared to the model with a free  $\mu$  parameter, in which the inferred mutation rate is  $\mu \approx 3 \cdot 10^{-6}$ . Models with  $\mu=10^{-5}$  and  $\mu=10^{-7}$  inferred different parameters estimates, including higher aneuploidy rate and lower fitness values for  $2n+1$  and  $2n^*$  (Figure S7).

**Extended informative prior distribution.** In an extended informative prior distribution, we used additional growth curves of  $2n^*$  (*refined* strain from Yona et al. (2012)) and  $2n+1$  in 39 °C to estimate  $w_{2n^*}/w_{2n+1}$  (Figure S3L). The same distribution was used for  $w_{2n^*}/w_{2n+1*}$ . Thus, our main informative prior uses a single prior distribution for fitness values of  $2n+1$ ,  $2n+1^*$ , and  $2n^*$ , whereas the extended informative prior uses one distribution for  $2n+1$ , and another distribution for both  $2n+1^*$  and  $2n^*$ .

We estimated the parameters under this extended informative prior. Inference took much longer to run but the posterior distribution seemed to converge, as it did not change much in the final iterations. The posterior predictive plot shows that inference with this extended prior produces a posterior distribution that fails to explain the empirical observations (pink in Figure 3). However, the inferred posterior distribution is considerably narrower (compare Figures 2 and S8) and therefore parameter estimates are less variable. The estimated mutation rate was much lower compared to the main informative prior, with  $\mu = 2.475 \cdot 10^{-9}$  [ $2.419 \cdot 10^{-9} - 2.609 \cdot 10^{-9}$ ]. Other parameter estimates are:  $\delta = 2.707 \cdot 10^{-3}$  [ $2.093 \cdot 10^{-3} - 3.092 \cdot 10^{-3}$ ],  $w_{2n+1} = 1.022$  [ $1.021 - 1.024$ ],  $w_{2n+1*} = 1.029$  [ $1.027 - 1.03$ ],  $w_{2n^*} = 1.03$  [ $1.029 - 1.031$ ]. Notably, the maximum posterior ratio  $w_{2n^*}/w_{2n+1} = 1.007$  is much lower than the maximum prior ratio of 1.033 (Figure S3H) and closer to the ratio of 1 that we assume in our standard prior. Together with the posterior predictive results, we conclude that the main informative prior is preferable over the extended informative prior.

## Supplementary Figures & Tables

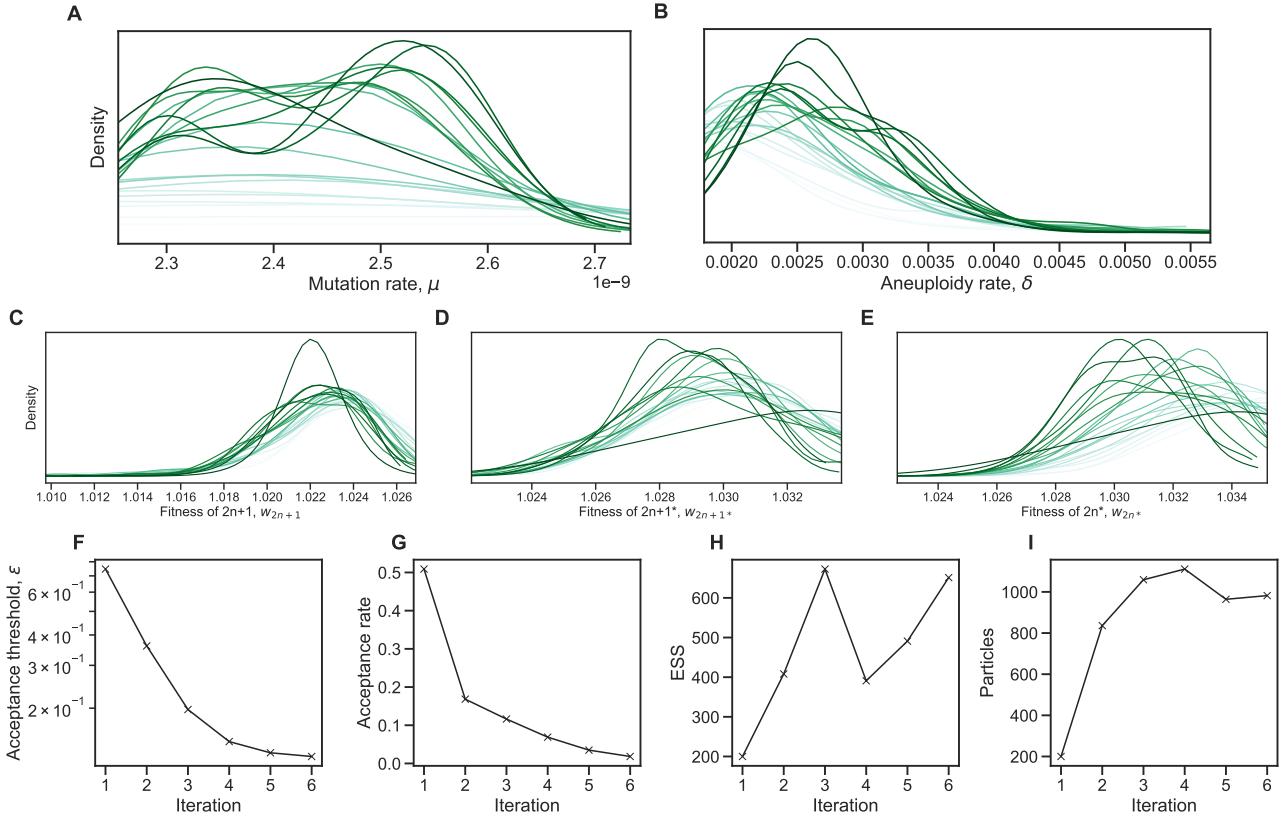


**Figure S1: Posterior distribution validation.** The posterior distribution of model parameters is roughly the same regardless of the number of simulations (4-10,000 replicates) used to approximate the likelihood (eq. (4)).

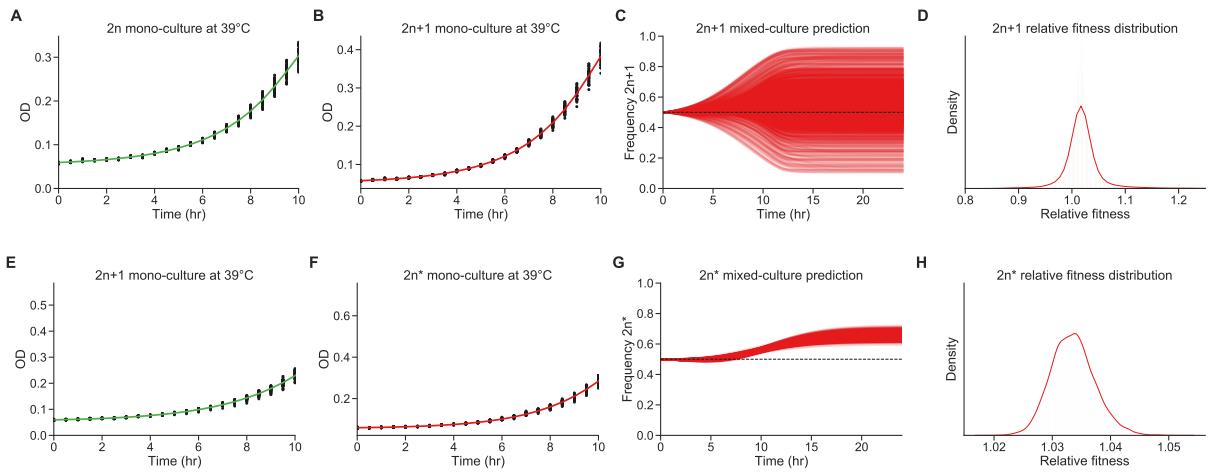
**Table S1: WAIC values for different  $\tau$  values.**

Model	WAIC
$\tau = 1$	295
$\tau = 33/32$	266
$\tau = 2$	501
$\tau = 5$	376
$\tau = 10$	318
$\tau = 100$	319

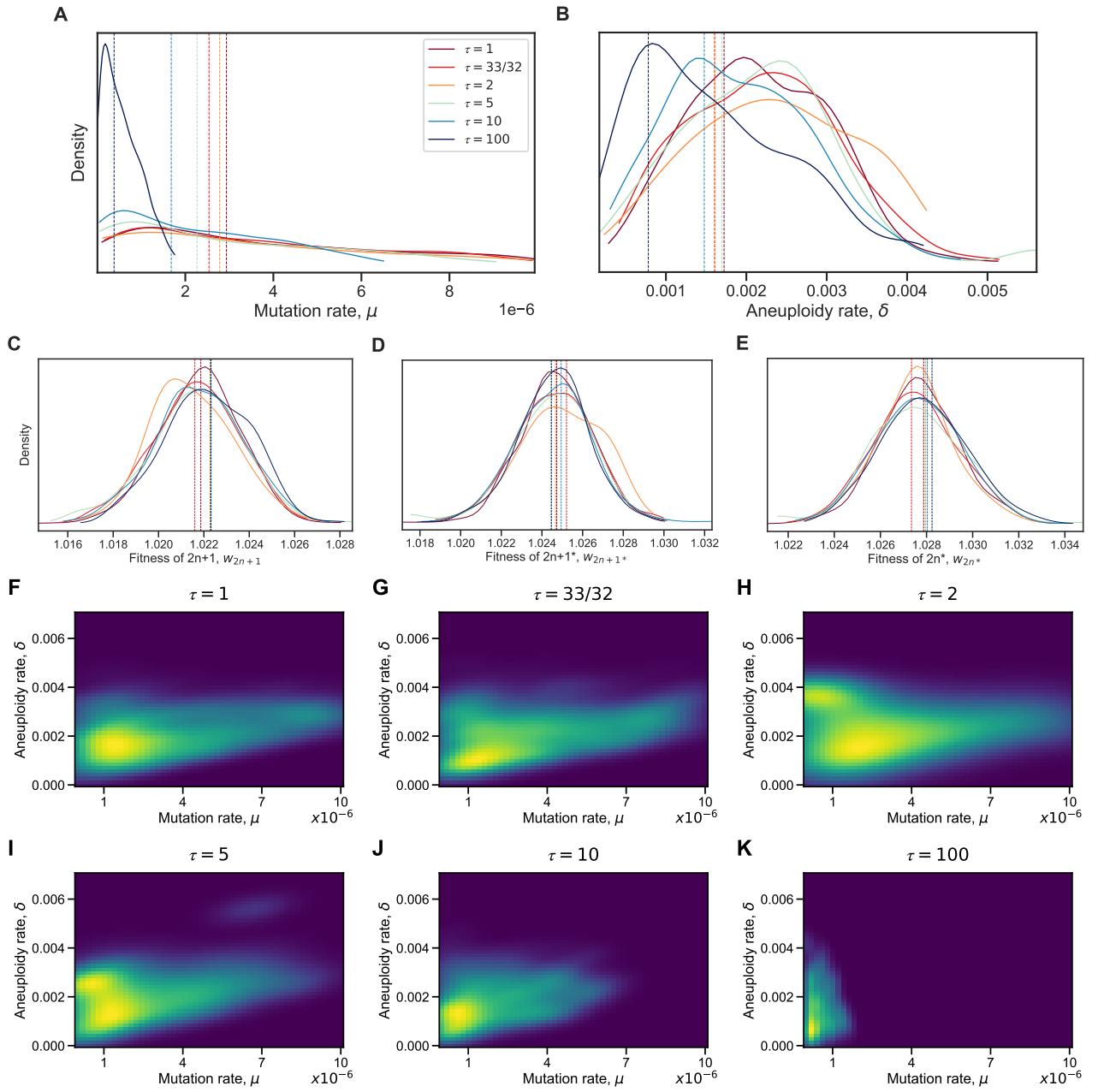
WAIC defined in eq. (6).



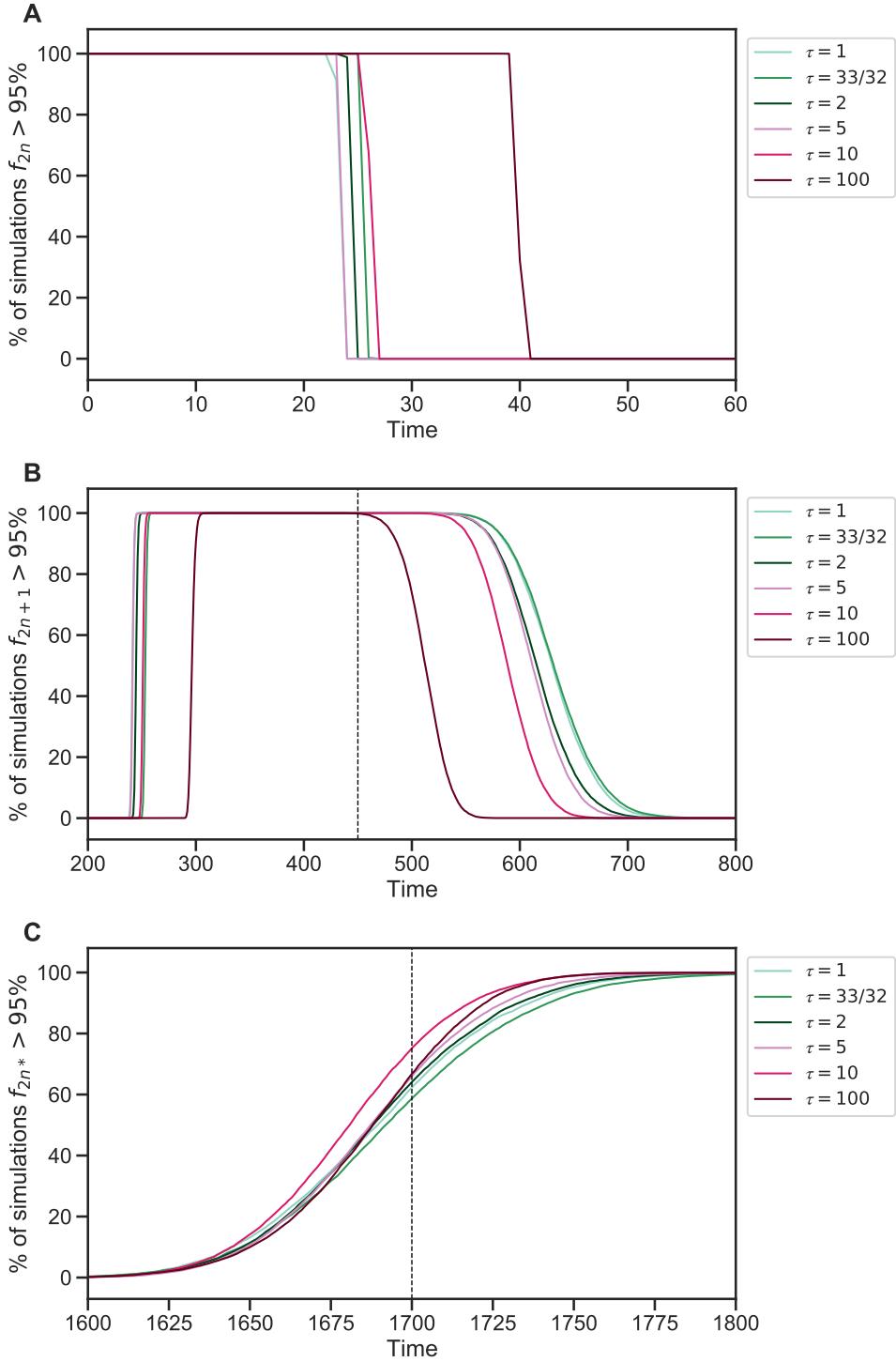
**Figure S2: Inference convergence.** The ABC-SMC algorithm was used to infer the model parameters. **(A-E)** The approximate posterior distributions of model parameters at each iteration of the ABC-SMC algorithm demonstrates convergence, as the posterior did not significantly change after the first iteration,  $t = 1$ . **(F-I)** ABC-SMC measures of convergence. After iteration number 6, the acceptance threshold was  $\epsilon = 0.13$  (i.e.,  $\mathcal{L} = 0.87$ , eq. (4)), the acceptance rate was 0.018, the number of particles was 982, and the effective sample size ESS=651.



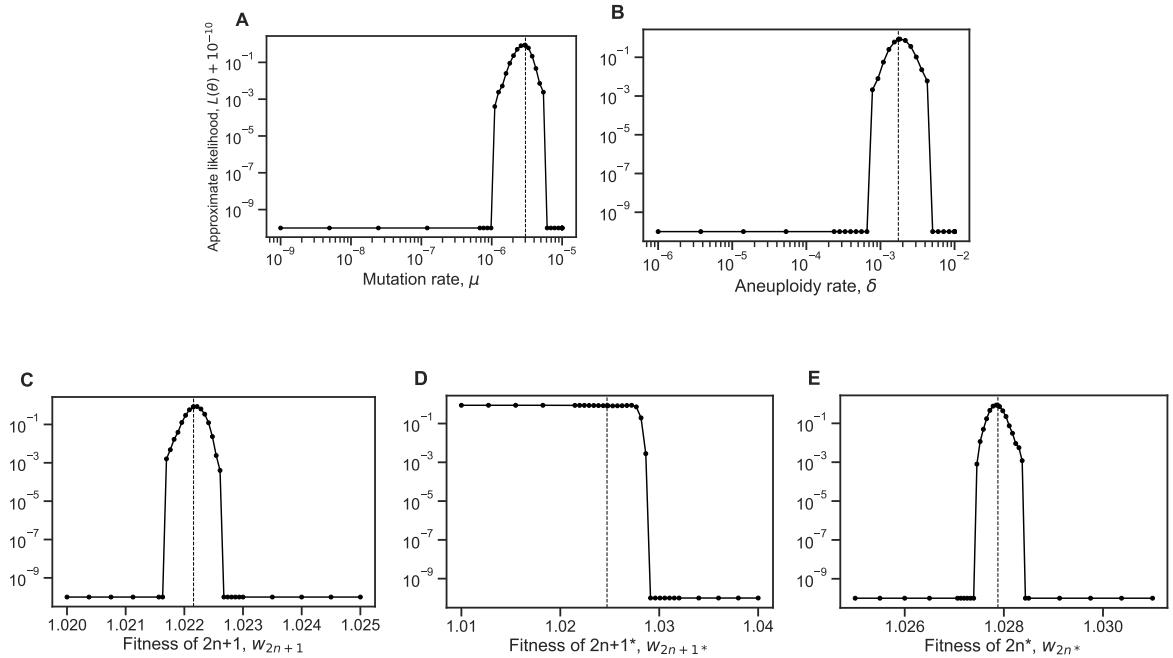
**Figure S3: Fitness estimation from growth curves.** (A-D) Fitness estimation from growth curves of  $2n$  and  $2n+1$  at  $39^{\circ}\text{C}$ .  $w_{2n+1}/w_{2n}=1.024$  (95% CI: 0.959 - 1.115). Curveball (E-H) Fitness estimation from growth curves of  $2n+1$  and  $2n^*$  at  $39^{\circ}\text{C}$ .  $w_{2n^*}/w_{2n+1}=1.033$  (95% CI: 1.027 - 1.041). Growth curves previously described in Yona et al. (2012, Figs. 3C, 4A, and S2). Fitness estimated from growth curves using Curveball, a method for predicting results of competition experiments from growth curve data (Ram et al., 2019, [curveball.yoavram.com](http://curveball.yoavram.com)). See *Models and Methods, Prior distributions* for more details. (A,B;E,F) Mono-culture growth curve data (markers) and best-fit growth models (lines). (C,G) The mixed-culture prediction for the strains from A,B and E,F respectively, 6,375 generated curves. (D,H) The relative fitness distribution for  $2n+1$  relative to  $2n$  (panel D) and  $2n^*$  relative to  $2n+1$  (panel H). Figures generated by Curveball.



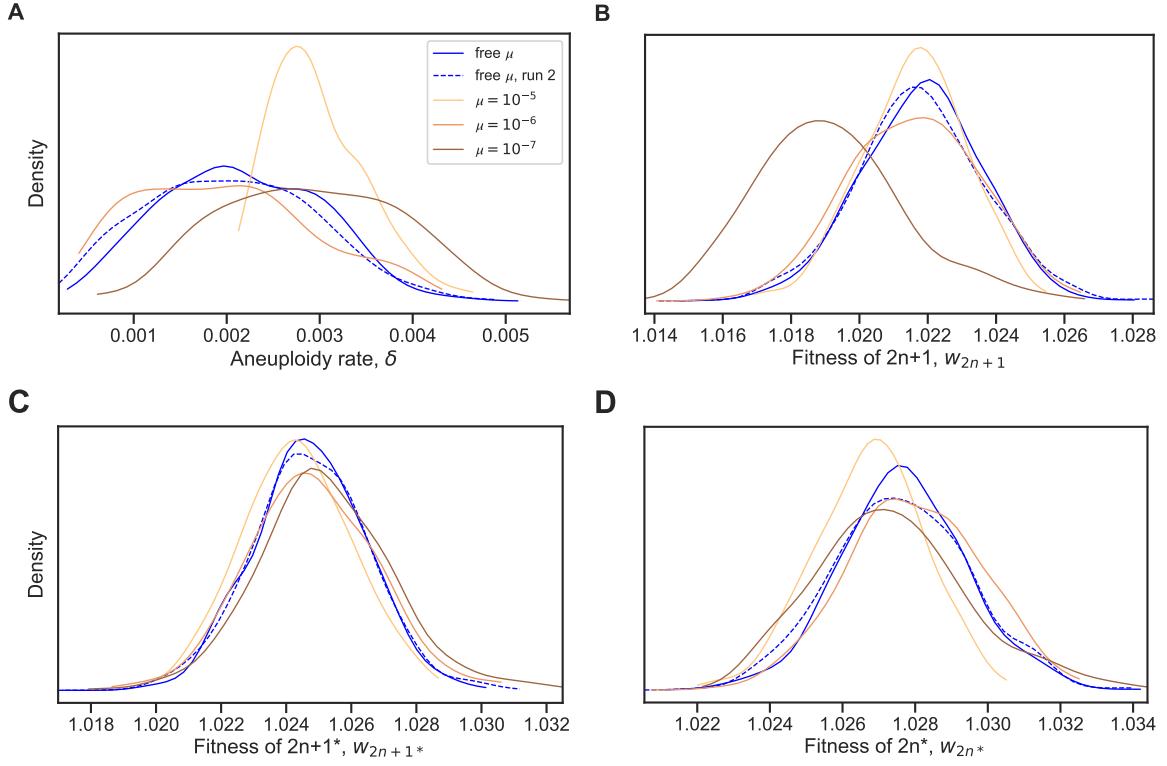
**Figure S4: Model with elevated mutation rate in aneuploid cells.** (A-E) The inferred posterior distributions for models with different values of  $\tau$ , the fold-increase in mutation rate in aneuploid cells ( $2n+1$  and  $2n+1^*$ ). Vertical dashed lines represent the MAP (maximum a posteriori) of each distribution. When the increase in mutation rate is high,  $\tau = 10$  and  $\tau = 100$ , the inferred mutation (A) and aneuploidy (B) rates tend to be lower. (F-K) The inferred joint posterior distribution of mutation rate ( $\mu$ ) and aneuploidy rate ( $\delta$ ) with different  $\tau$  values (dark purple and bright yellow for low and high density, respectively).



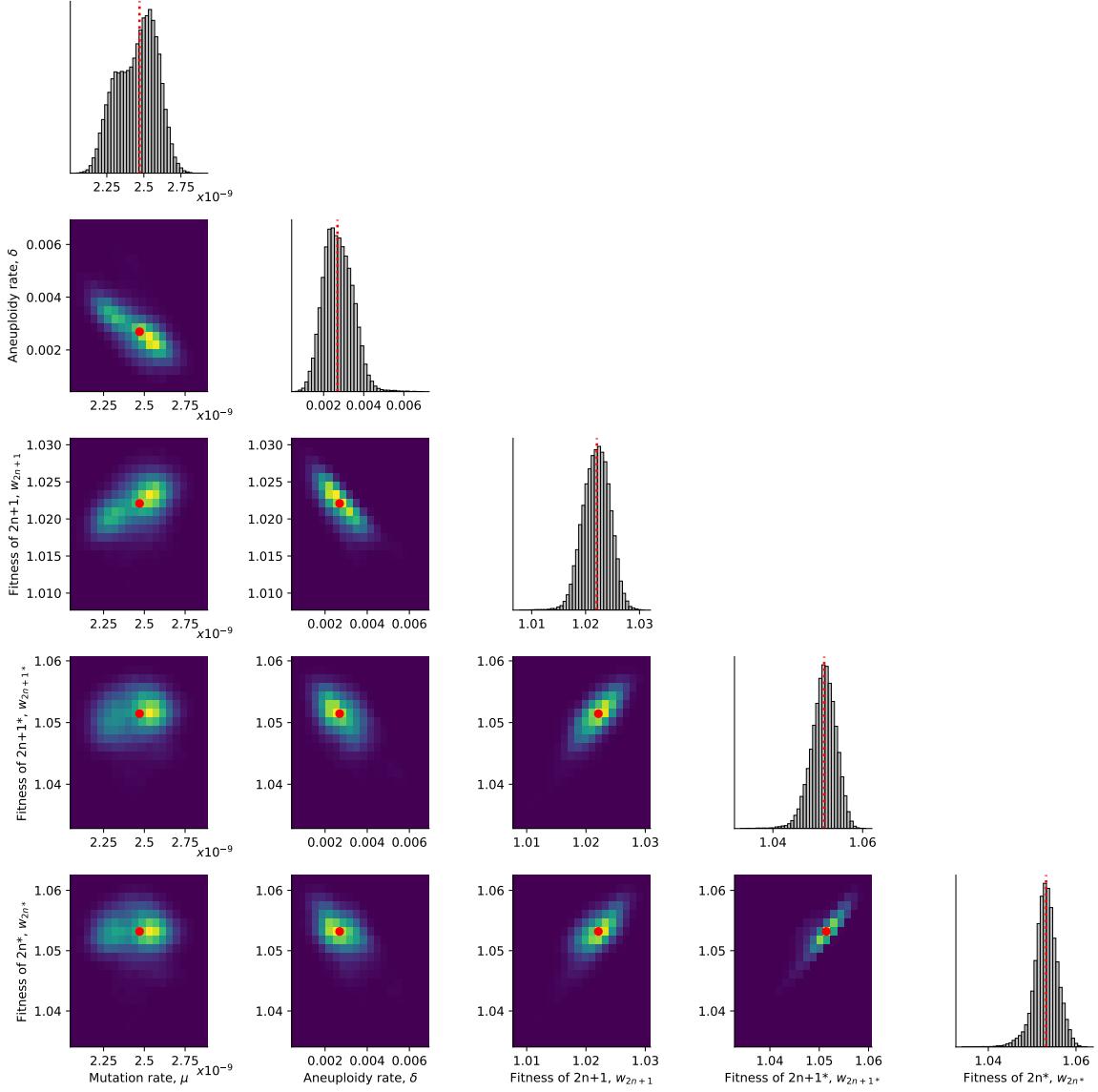
**Figure S5: Genotype fixations for models with increased genetic instability.** We estimated the parameters for different models, each assuming a different value of  $\tau$ , the fold-increase in mutation rate in aneuploid cells. We then generated 10,000 simulations using the MAP estimate of each model and evaluated the fraction of simulations in which the frequency of genotype  $2n$  (**A**),  $2n+1$  (**B**), and  $2n^*$  (**C**) is above 95% (y-axis) at each generation (x-axis). Note that  $2n+1^*$  did not fix. We can see that  $\tau = 100$  can be distinguished if the waiting time for  $f_{2n} < 95\%$  is known (panel A) or if the waiting time for  $f_{2n+1} > 95\%$  or  $f_{2n+1} < 95\%$  is known (panel B). It is harder to distinguish between  $1 \leq \tau \leq 10$ .



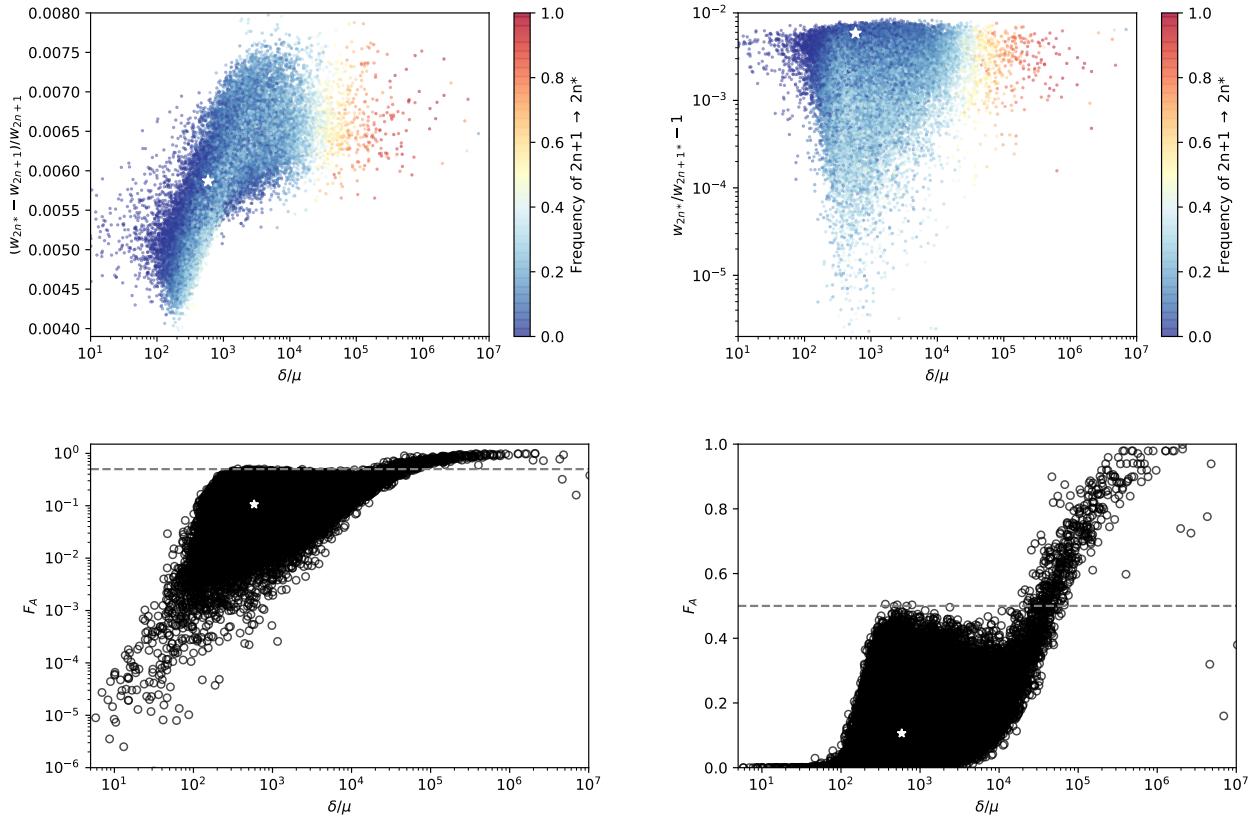
**Figure S6: Likelihood profiles.** Sensitivity of the model approximate likelihood,  $\mathcal{L}(\theta)$ , to changing a single parameter while the other parameters remain fixed at their MAP estimates. Dashed vertical line represents the MAP value. The prior distributions for the mutation rate and aneuploidy rate are  $\mu \sim U(10^{-9}, 10^{-5})$  and  $\delta \sim U(10^{-6}, 10^{-2})$ , respectively.



**Figure S7: Model with fixed mutation rate.** (A-D) The inferred posterior distributions for models with free and fixed mutation rate,  $\mu$ . The MAP (maximum a posteriori) and 50% HDI (highest density interval) for each model are: **free  $\mu$ , run 1:**  $\delta = 2.749 \cdot 10^{-3}$  [ $1.476 \cdot 10^{-3} - 2.822 \cdot 10^{-3}$ ],  $w_{2n+1} = 1.022$  [1.021–1.023],  $w_{2n+1^*} = 1.025$  [1.023–1.026],  $w_{2n^*} = 1.027$  [1.026–1.029]; **free  $\mu$ , run 2:**  $\delta = 1.938 \cdot 10^{-3}$  [ $1.338 \cdot 10^{-3} - 2.748 \cdot 10^{-3}$ ],  $w_{2n+1} = 1.022$  [1.02–1.023],  $w_{2n+1^*} = 1.025$  [1.023–1.026],  $w_{2n^*} = 1.027$  [1.026–1.029];  **$\mu = 10^{-5}$ :**  $\delta = 3.089 \cdot 10^{-3}$  [ $2.412 \cdot 10^{-3} - 3.169 \cdot 10^{-3}$ ],  $w_{2n+1} = 1.022$  [1.021–1.023],  $w_{2n+1^*} = 1.024$  [1.023–1.026],  $w_{2n^*} = 1.027$  [1.026–1.028];  **$\mu = 10^{-6}$ :**  $\delta = 1.413 \cdot 10^{-3}$  [ $1.04 \cdot 10^{-3} - 2.529 \cdot 10^{-3}$ ],  $w_{2n+1} = 1.021$  [1.02–1.023],  $w_{2n+1^*} = 1.024$  [1.023–1.026],  $w_{2n^*} = 1.028$  [1.026–1.029];  **$\mu = 10^{-7}$ :**  $\delta = 3.4 \cdot 10^{-3}$  [ $2.043 \cdot 10^{-3} - 3.578 \cdot 10^{-3}$ ],  $w_{2n+1} = 1.019$  [1.017–1.02],  $w_{2n+1^*} = 1.026$  [1.024–1.027],  $w_{2n^*} = 1.027$  [1.026–1.029].



**Figure S8: Posterior distribution of parameters inferred with the extended prior distribution.** On the diagonal, the inferred posterior distribution of each model parameter. Below the diagonal, the inferred joint posterior distribution of pairs of model parameters (dark purple and bright yellow for low and high density, respectively). Red markers and orange lines for the joint MAP estimate (which may differ from the marginal MAP, as the marginal distribution integrates over all other parameters).



**Figure S9: Posterior distribution of  $F_A$ .** (A,B)  $F_A$  values (color coded) as in Figure 4 for different parameter choices on the x- and y-axes. White start denotes the MAP estimate. (C,D)  $F_A$  values (y-axis) as a function of  $\delta/\mu$ , the aneuploidy-to-mutation rate ratio.