A fitness trade-off explains the early fate of yeast aneuploids with chromosome gains

Simone Pompei[a] and Marco Cosentino Lagomarsino[a,b,c,1]

The early development of aneuploidy from an accidental chromosome missegregation shows contrasting effects. On the one hand, it is associated with significant cellular stress and decreased fitness. On the other hand, it often carries a beneficial effect and provides a quick (but typically transient) solution to external stress. These apparently controversial trends emerge in several experimental contexts, particularly in the presence of duplicated chromosomes. However, we lack a mathematical evolutionary modeling framework that comprehensively captures these trends from the mutational dynamics and the trade-offs involved in the early stages of aneuploidy. Here, focusing on chromosome gains, we address this point by introducing a fitness model where a fitness cost of chromosome duplications is contrasted by a fitness advantage from the dosage of specific genes. The model successfully captures the experimentally measured probability of emergence of extra chromosomes in a laboratory evolution setup. Additionally, using phenotypic data collected in rich media, we explored the fitness landscape, finding evidence supporting the existence of a per-gene cost of extra chromosomes. Finally, we show that the substitution dynamics of our model, evaluated in the empirical fitness landscape, explains the relative abundance of duplicated chromosomes observed in yeast population genomics data. These findings lay a firm framework for the understanding of the establishment of newly duplicated chromosomes, providing testable quantitative predictions for future observations.

aneuploidy | evolutionary modeling | yeast evolution | cancer evolution

Aneuploidy, a deviation from the normal chromosome number, is a form of large-scale genomic variation, involving changes both at the genotypic level and at the phenotypic level, and one of the hallmarks of cancer (1). In cancer genomes, aneuploidy correlates with important genomic changes, such as TP53 mutation and expression of proliferation genes (2) and drug resistance mutations leading to treatment failure (3, 4). Drugs that disrupt mitotic progression, called antimitotic drugs (5, 6), are widely used for cancer treatment. These drugs cause chromosome missegregation, large genetic rearrangements, and aneuploidy. In *Saccharomyces cerevisiae* models (hereafter referred to as yeast), perturbed gene expression due to extra chromosomes can cause stress resulting from the proteome-wide stoichiometric imbalance of protein levels (7, 8). Moreover, aneuploidy was shown to cause global changes in mRNA and protein expression and to possibly confer condition-dependents fitness advantage (9–11). Finally, in yeast, the ploidy levels were shown to influence the emergence of mutator strains (12), to alter the effect of genetic mutations (13), to influence the phenotypic attributes of hybrid individuals (14), to impact the speed of adaptation (15, 16), and to act on transcriptional silencing (17) and on pleiotropic effects (18).

The evolutionary dynamics leading to the emergence of aneuploidy is typically investigated with yeast models because they can be manipulated with advanced genetics and cell- and molecular-biology methods and hence be used to create isogenic backgrounds that differ from each other only by chromosome ploidy number, offering a direct point of comparison between euploid and aneuploid strains. Moreover, yeast models can be easily investigated with laboratory-evolution experiments, thanks to their short replication time (6, 9, 19, 20).

Several experiments in the last decades have raised apparently controversial evidence for the evolutionary role of aneuploidy (10, 21). While aneuploidy carries significant cellular stress and decreased fitness (22)—measured, for example, by a reduction of growth rate—it has also been shown to carry a beneficial effect that provides a quick, transient solution to external stress (20, 21). This quick solution often emerges faster, hence more frequently, than other evolutionary routes. Intriguingly, cultured human cells show the same contrasting trends: Aneuploid cancer cells lines show a reduction of

[1]To whom correspondence may be addressed. Email: marco.cosentino-lagomarsino@ifom.eu.

growth rate (23–25), but specific patterns of aneuploidy, particularly in the presence of extra chromosomes, confer a beneficial effect in specific adverse conditions (26, 27).

In the case of yeast, the literature offers extensive phenotypic data, for example, growth curves of aneuploids in several environmental conditions as well as in laboratory-evolution experiments (6, 9, 10, 19–21), offering the opportunity to test for unifying trends. The available modeling studies presented so far have focused on the effect of aneuploidy on cell growth and physiology (7, 28–31). Two interesting recent studies (28, 30) proposed a stochastic model of evolution similar to the classic Fisher's geometrical model (32) to describe the fitness landscape of a set of engineered aneuploid strains in different stress environments. This model explains the observed correlation between the degree of phenotypic variation and the degree of overall growth suppression, measured in ref. 9. However, this model and all the approaches presented so far (7, 28–31) are limited by a static description of the genetic and phenotypic architecture of aneuploids, failing to provide a description of the mutational dynamics leading to its emergence.

Here, we develop a theoretical framework to describe such mutational dynamics and to address the cost–benefit trade-offs in early aneuploids. We introduce a fitness model where a fitness cost proportional to the number of genes in the duplicated chromosome is counterbalanced by a fitness advantage resulting from the dosage increase of specific genes. Our approach builds on the so-called "mutation bias" framework (33–35), a class of evolutionary models used to investigate the role of fast mutational processes in directing evolution, in a scenario where evolutionary routes emerging with a high mutation rate are in evolutionary competition with alternative mutational targets generated with a lower rate but able to confer a higher fitness advantage. Our model makes quantitative predictions that capture the dynamics leading to the emergence of aneuploidy. As we will describe in detail, the model captures the probability of the emergence of extra chromosomes in experimental setups (20) and correctly predicts the observed outcomes for the emergence of aneuploidy. We then make use of phenotypic data to isolate the main features contributing to the fitness landscape of aneuploids with extra chromosomes and show that the dynamics of our model in this landscape captures the relative abundance of aneuploidies observed in population genomics data.

Results

Model and Parameters. We develop and analyze an evolutionary model to describe the emergence of aneuploidy carrying extra chromosomes. Fig. 1*A* describes the key model ingredients. The model considers a population consisting of euploid individuals, which is exposed to an external stress causing a decrease in their growth rate. Individuals in the population can respond to the stress by increasing the expression of a specific target gene, gaining a beneficial effect quantified by the selection coefficient ($\sigma_b > 0$). Individuals can gain fitness by two alternative evolutionary routes: i) by increasing the target gene expression (for example, with mutations on the promoter binding site) or improving its functionality via a set of point mutations (on coding regions adapting protein function), occurring at a total rate μ_m or ii) via missegregation events, taking place at a higher rate ($\mu_a > \mu_m$) and resulting in the emergence of aneuploid individuals carrying extra chromosomes. Note that route (i) could require several point mutations (modeled here as a one-step process), but the per-base mutation rate is a lower bound for μ_m. Aneuploids with extra chromosomes are less fit than euploids because the duplication

of the nontarget genes in the extra chromosome determines a global fitness cost ($\sigma_c > 0$). Hence, the selection coefficient of aneuploids ($\sigma_b - \sigma_c$) is lower than that of euploid mutants (σ_b), and euploid mutants, although generated at a lower rate, have a higher fixation probability than aneuploids ($\phi_m > \phi_a$). Double mutants (individuals carrying both aneuploidy and point mutations) are produced at a rate corresponding to the product of the rates ($\mu_m \times \mu_a$) and therefore are very rare and can be neglected. We also assume that all mutations other than missegregations and the target point mutations do not contribute to the adaptive dynamics in response to the external stress; hence, we neglected them. Under these assumptions, the model reduces to the competition between two possible beneficial mutations, aneuploidy vs. a local mutation.

Importantly, in our model, we do not consider an explicit fitness landscape, connecting the increase of gene expression and fitness but we consider only a simplified model of the selection coefficients of the two mutants. More specifically, we focus on a scenario where point mutations induce an increase of the expression of the target gene which is similar to that of the aneuploid individuals (i.e., $2x$ for ploidy = 1 background and $1.5x$ for the ploidy = 2 background). In other words, the benefit of the aneuploidy would have the same selection coefficient (σ_b) as the euploid beneficial mutation.

However, in the more general scenario, point mutations might alter gene expression to different degrees and are likely to cause an increase of the expression which is lower than the effect of gene duplication (36). In this case, the two selection coefficients (aneuploidy and euploid mutant) would differ, and this difference would depend on the properties of the fitness landscape. In the rather general case where there is an optimal value of the gene expression that corresponds to a fitness peak (37, 38), the fitness landscape could be described by a quadratic function (39), and the mutant with the highest selection coefficient would be the one whose expression is closer to the optimal value.

If the optimal value is close to the expression resulting from a chromosome duplication (i.e., $2x$ for ploidy = 1 background and $1.5x$ for the ploidy = 2 background), then any point mutation causing a lower or a higher variation of expression would have a lower selection coefficient. The reduced selection coefficient would result in a lower fixation probability ϕ_m and in a lower fixation rate. We note that this more general model can be effectively described by our framework by a reduction of the mutation rate $\mu_m^{\text{eff}} < \mu_m$, while keeping the same fixation probability (and hence, the same selection coefficient). In other words, for the fixation dynamics, having a lower selection coefficient corresponds to having a lower mutation rate, allowing us to map on our model scenarios where point mutations of euploid individuals alter the expression of the target gene to a different degree than a full chromosome duplication. In a similar way, the case where the optimal value of the fitness landscape is close to the gene expression resulting from point mutations would be effectively described with an increase of the mutation rate of the euploid mutant $\mu_m^{\text{eff}} > \mu_m$, while keeping the same selection coefficient of the aneuploid mutant. (This point is further explained in *SI Appendix*.)

Evolutionary Dynamics. Our question concerns the conditions in which chromosomal duplications emerge first. Accordingly, we investigate the "early-stage" population defined by the point in time when one of the two mutants, the euploid with point mutations or the aneuploid mutant carrying extra chromosomes, becomes fixed in the population (i.e., reaches an intrapopulation
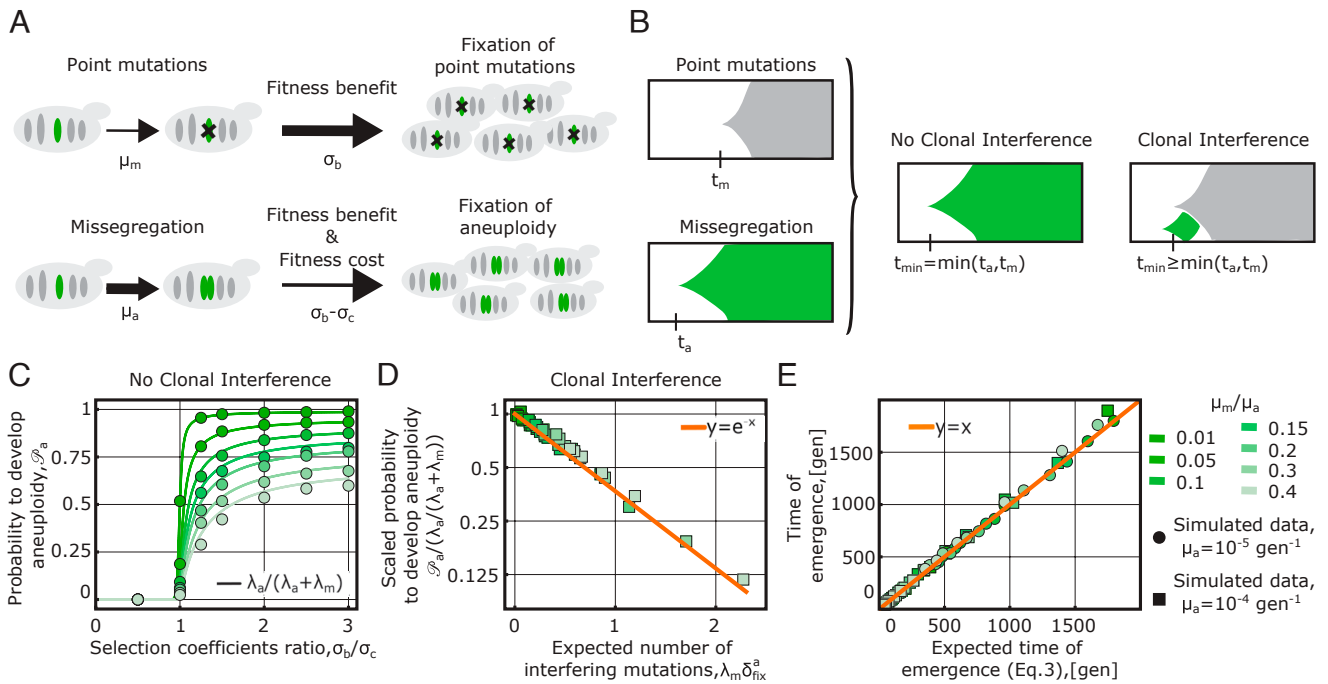
Fig. 1. A trade-off between fitness cost and fitness benefit explains the population dynamics of early aneuploids with extra chromosomes. (*A*) Schematic illustration of the evolutionary model, which considers two alternative mutational routes to cope with an external stress requiring the increase in the expression of a specific target gene. Individuals in an evolving population can increase the target gene expression via point mutations, taking place at rate (μ_m), or via duplication of the target chromosome, at a rate related to missegregation. This second evolutionary route takes place at a higher rate ($\mu_a > \mu_m$) and generates aneuploids. Aneuploids with extra chromosomes pay a fitness cost because they carry the duplication of nontarget genes in the extra copy of the chromosome containing the target gene. Hence, euploids, although emerging at a lower rate, have a higher fixation probability than aneuploids $\phi_m > \phi_a$. (*B*) The dynamics associated to individuals carrying mutations associated to the two evolutionary routes, i.e., aneuploid mutants (green) versus euploid mutants (grey) here schematically represented with Müller plots, which are used to illustrate the succession of genotypes in an evolutionary process (the horizontal axis shows time, while the vertical axis represents relative abundances of genotypes. Each genotype is shown with shaded areas of different colors and originates in an arbitrary clone placed in the middle of its parent area) is characterized by the time of emergence of the successful mutant (the mutant whose descendants will eventually take over the population). In the model, both evolutionary routes are attainable, and only the fastest of the two mutants, (i.e., the one that will generate a mutation able to overcome the genetic drift and reach fixation), emerging at a time $t_{\min} = \min(t_a, t_m)$, will reach fixation. Clonal interference—occurring, for example, when a euploid mutant emerges during the fixation dynamics of an aneuploid individual—will prevent the fixation of aneuploidy and effectively delay the emergence of the successful mutant ($t_{\min} \geq \min(t_a, t_m)$). (*C*) Fixation probability of aneuploids with extra chromosomes in a regime with no clonal interference ($\lambda_m \delta_{\text{fix}}^a \simeq 0$), plotted as a function of the nondimensional ratios σ_b/σ_c (x axis) and μ_m/μ_a (color coded). Results of simulations (circles) are compared to analytical calculations (solid lines). (*D* and *E*) Collapse plots of simulated data of the model in the clonal interference regime ($\lambda_m \delta_{\text{fix}}^a \geq 0$) validate the analytical results for the fixation probability of aneuploids with extra chromosomes (Eq. **1**, shown in panel *D*) and for the emergence time of the successful mutant (Eq. **3**, shown in panel *E*). *Material and Methods* for details on the numerical simulations of the model. Additional model parameters used for the data shown in panels *B*, *C*, and *D*: $N = 1000$, $\sigma_c N = 50$.

frequency $\simeq 1$) for the first time. This dynamics is described by fixation rates, which are given by the product of the mutation rates, the effective population size N, and the fixation rates: $\lambda_m = \mu_m N \phi_m(\sigma_m, N)$ for the euploid mutant and $\lambda_a = \mu_a N \phi_a(\sigma_a, N)$ for the aneuploid one. The fixation probabilities depend on the selection coefficients ($\sigma_a \equiv \sigma_b - \sigma_c$ for the aneuploid and $\sigma_m \equiv \sigma_b$ for the euploid mutant) and on the effective population size N, as given by Kimura's formula $\phi(\sigma, N) = (1 - e^{-2\sigma})/(1 - e^{-2\sigma N})$ (40).

Analytical Expression of the Probability to Develop Aneuploidy with Extra Chromosomes. In order to characterize the onset of the fastest variant, we focus on the waiting times for the emergence of a successful mutant, defined as the mutant that will eventually reach fixation. The two times, denoted as t_a and t_m for the aneuploid carrying extra chromosomes and the euploid mutant respectively, are stochastic variables. Since these mutations emerge at a constant rate, the probability distribution of the waiting times is exponential $t_{a,m} \sim Exp(\lambda_{a,m})$; hence, their expected values are equal to the inverse of the fixation rates ($\tau_{a,m} \equiv \langle t_{a,m} \rangle = 1/\lambda_{a,m}$).

The statistics of the fastest emerging mutant can be described by the difference of the two times, $t_{\text{diff}} \equiv t_a - t_m$, whose

probability density has an analytical expression (*SI Appendix*). In particular, the problem of computing the probability for the variant carrying extra chromosomes to reach fixation is equivalent to computing the probability for the time difference to be negative ($t_{\text{diff}} < 0$). However, since the selection coefficient of aneuploids carrying extra chromosomes is lower than that of the euploid mutant, individuals of the former class will interfere with the expected progression of the aneuploid mutation to the fixation (Fig. 1*B*), by an effect known as "clonal interference" (CI) (41–44).

We find that, to compute the probability to fix extra chromosomes, CI effects are captured by the extended condition $t_{\text{diff}} + \delta_a^{\text{fix}} < 0$, where $\delta_a^{\text{fix}} = \log(2N\sigma_a)/\sigma_a$ is the effective time to fixation of an aneuploid mutant carrying extra chromosomes (44) (i.e., the time interval during which CI effects can take place, *Material and Methods*). This leads to the expression

$$\mathcal{P}_a \equiv P(t_{\text{diff}} + \delta_a^{\text{fix}} < 0) = \frac{\lambda_a}{\lambda_m + \lambda_a} e^{-\lambda_m \delta_a^{\text{fix}}}, \qquad [1]$$

for the fixation probability. This expression is similar to the ones presented in refs. 35 and 45. Consistently with ref. 41, CI effects are related to the expected number of euploid mutations that can

emerge during the fixation dynamics of the mutant with extra chromosomes. In the limit $\lambda_m \delta_a^{\text{fix}} \ll 1$, there are no interfering mutations, and the probability to develop extra chromosomes is set by the fixation rates alone $\mathcal{P}_a \simeq \frac{\lambda_a}{\lambda_m + \lambda_a}$ (Fig. 1C). In the clonal interference regime $\lambda_m \delta_a^{\text{fix}} > 1$, the emergence of aneuploidy with extra chromosomes is exponentially suppressed to zero ($\mathcal{P}_a \propto e^{-\lambda_m \delta_a^{\text{fix}}}$, Fig. 1D). Moreover, in this regime, the evolutionary dynamics would be characterized by the elimination of aneuploidy, resulting from the emergence of euploid mutants, after an initial increase in the frequency of aneuploid mutants. However, the observed loss of anueploidy would not signal the existence of karyotype instability, as the duplicated chromosome would not be lost within aneuploid individuals. Our model can be exploited to investigate this scenario and predicts the existence of a critical population size around which such dynamics (rise of aneuploidy to high frequency and subsequent elimination because of CI effects) could be observed (*SI Appendix*, Fig. S9).

When $\mathcal{P}_a > \frac{1}{2}$, the emergence of aneuploidy is more likely than that of the competing beneficial point mutations. Hence, the condition $\mathcal{P}_a = \frac{1}{2}$ sets a lower critical value for the beneficial selection coefficient (σ_b^*), which reads (*SI Appendix* for derivation)

$$\sigma_b^* = \frac{\sigma_c}{1 - r} + \mathcal{O}(\mu_m N \log N). \qquad [2]$$

The above equation defines σ_b^*. The prevalence of the evolutionary route developing extra chromosomes is observed in "stress" conditions where the beneficial effect exceeds the minimal value $\sigma_b \geq \sigma_b^* > \sigma_c$. Here, $r = \mu_m / \mu_a < 1$ is the ratio between the mutation μ_m and the missegregation rate μ_a (*SI Appendix*).

Aneuploidy with Extra Chromosomes Is a "Quick Fix" in Stressful Conditions. The dynamics leading to the fixation of one of the two evolutionary routes can also be described in terms of the waiting time before the emergence of the fastest successful mutant. This dynamical quantity is described by the minimum of the two waiting times, $t_{\min} \equiv \min(t_a, t_m)$, and has expected value (Fig. 1E and *SI Appendix* for derivation)

$$\tau_{\min} = \langle t_{\min} \rangle = \tau_m \left(1 - (1 + \lambda_m \delta_{\text{fix}}^a) \mathcal{P}_a \right). \qquad [3]$$

Thanks to the possibility of developing extra chromosomes, the waiting time until the emergence of the successful mutant is therefore shorter than the time needed to develop the competing set of point mutations $\tau_m = 1/\lambda_m$, which, in our model, would be attained if the mutational route was the only genomic change offering a solution to the external stress. This evolutionary route is still dynamically selected when $\sigma_b < \sigma_c \rightarrow \tau_{\min} \simeq \tau_m = 1/\lambda_m$, i.e., when the global effect of extra chromosomes (benefit minus cost) is detrimental. Conversely, in the opposite limit $\sigma_b \gg \sigma_c \rightarrow \tau_{\min} \simeq \tau_a = 1/\lambda_a$, the waiting time is set up by the fixation rate of extra chromosomes alone. Clonal interference effects ($\lambda_m \delta_{\text{fix}}^a > 0$) lead to an increase of the waiting time, i.e., reducing the speed of adaptation in response to the stress, consistently with refs. 41–44.

In summary, the model describes in quantitative terms the early-stage evolutionary role of aneuploidy carrying extra chromosomes. According to the predictions, extra chromosomes provide a "quick fix" to the external stress (because $\mathcal{P}_a \simeq 1 \rightarrow \tau_{\min} \simeq \tau_a < \tau_m$). Aneuploidy also has an indirect effect on the mutational dynamics of euploid individuals, by effectively selecting the fast mutants, hence causing a reduction of the waiting time to the emergence of the successful euploid mutant ($1 > \mathcal{P}_a > 0 \rightarrow \tau_{\min} < \tau_m$).

The Model Correctly Predicts the Outcome of Experimental Evolution Data from Ref. 20. Our model can be applied to describe the evolutionary dynamics observed in experimental setups akin to ref. 20. In their experiment, Yona and coworkers exposed four independent yeast populations of diploid strains to a constant heat stress of 39°C. After ~450 generations, the duplication of chromosome III (trisomy) was found to have reached fixation in all four populations. The duplication of this chromosome was shown to carry a beneficial effect in response to the applied heat stress and to be the dominant evolutionary solution over an alternative mutational route attained by point mutations inducing the upregulation of heat-shock genes.

In order to compare the model prediction to the outcome of the experiment, we obtained growth curves from the authors of ref. 20, evaluated for the diploid and aneuploid strain (carrying the trisomy of chromosome III) both in normal conditions (30 °C) and in stress conditions (39 °C). We used the growth curves to infer values of the selection coefficients of aneuploid individuals. (*Materials and Methods*, *SI Appendix*, Fig. S1 and Table S1; the numerical values we obtained are $\sigma_b = 0.17 \, \text{gen}^{-1}$ and $\sigma_c = 0.05 \, \text{gen}^{-1}$).

Given these values for the selection coefficients, we evaluated the cumulative probability of developing aneuploidy (Eq. **13**) vs time, according to our model prediction (Eqs. **1** and **3** and *Material and Methods*), using an effective population size of $N = 10^6$ individuals, as a function of the missegregation rate (μ_a) and of the total mutation rate (μ_m) (Fig. 2). We find the model predictions Eqs. **1** and **3** to be in quantitative agreement with the outcome of the experiment, for realistic values of the missegregation rate ($\mu_a \geq 8 * 10^{-7} \text{gen}^{-1}$) and of the total mutation rate ($\mu_m \leq 5 * 10^{-9}$). Similar results are obtained by setting a bigger value for the effective population size of ($N = 10^7$; *SI Appendix*, Fig. S3).

In order to determine realistic ranges of the rates (μ_a, μ_m) in yeast, we reasoned as follows. The numerical value of the yeast per-base mutation rate is $\mu_{\text{spont.}} = 1.7 * 10^{-10} \text{gen}^{-1}$ (46). The mutation rate can be higher than the spontaneous rate since the same phenotypic effect, i.e., the development of resistance to heat by upregulation of heat-shock genes, can be attained with more than a single point mutation. A conservative estimate of the size of this mutational target is no more than 100 bases, giving an upper bound constraint for $\mu_m \leq 10^{-8} \text{gen}^{-1}$. Values of the mutation rate lower than the spontaneous rate, on the other hand, would correspond to a scenario where the selection of the euploid mutant does not develop a whole duplication of the expression of the target gene (but would alter gene expression to a lower degree; *SI Appendix*) and hence were also considered to be realistic. Measurements for the missegregation rate exist in the literature ($\mu_a \simeq 10^{-6} \text{gen}^{-1}$).

The agreement between model prediction for the probability to develop aneuploidy and for the time fo fixation and experimental data is observed in yet another independent evolutionary experiment, described in refs. 20 and 47, where a diploid yeast population was exposed to a different stressing environment, high pH. This experiment revealed the fixation of strains with the duplication of chromosome V (trisomy; *SI Appendix*, Figs. S1, S2, and S3 and Table S1).
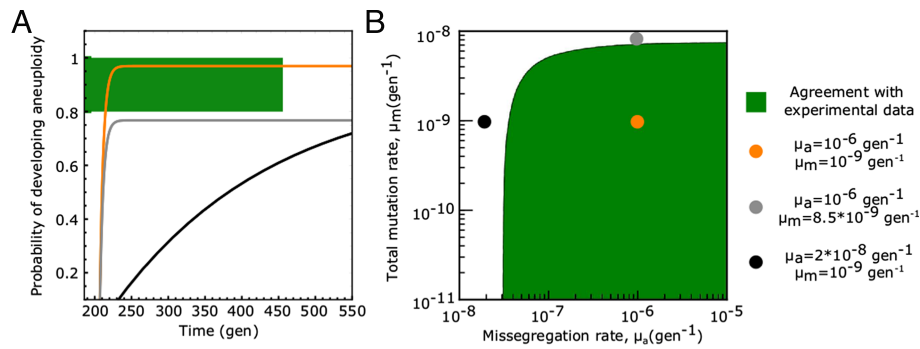
Fig. 2. Model predictions agree with laboratory-evolution data from ref. 20. (*A*) Expected cumulative probability for the emergence of aneuploidy with extra chromosomes vs. the time to reach fixation (*Material and Methods*), computed according to the model prediction (Eqs. **1** and **3**) shown for three combinations of the values of the model parameters (μ_a, μ_m) (color-coded, numerical values reported in the legend of the plot). In the experiment, where a yeast population was exposed to stress by increasing the temperature to 39 °C, 4 out of 4 yeast populations developed chromosomal duplications ($CI_{66\%} = [0.8, 1]$ for the probability to develop aneuploidy), and all the fixations were reached before 450 generations. Hence, the experimental data fall in the region of the plot corresponding to $P_a \in [0.8, 1]$ and $t = 450\,gen$, marked by a green bar and highlighted by green dashed lines. The trajectories predicted by the model that cross this region are in agreement with the experimental data. Similarly, panel (*B*) shows the combinations of the numerical values of the model parameters (μ_a, μ_m) that are in agreement with the experimental data. The colored circles mark the values of the model parameters that were used to generate the trajectories shown in *A* (each dot has the same color of the corresponding trajectory in *A*). Numerical values of the beneficial selection coefficient ($\sigma_b = 0.17\,\mathrm{gen}^{-1}$) and for the fitness cost of aneuploidy ($\sigma_c = 0.05\,\mathrm{gen}^{-1}$) were obtained from exponential fits of the growth curves of the corresponding yeast strains (20) (*Material and Methods* and *SI Appendix*, Fig. S1). The effective population size was set to $N = 10^6$ individuals (*SI Appendix*, Fig. S3 shows results for $N = 10^7$). The data reported here refer to the "high-temperature" experimental setup. Similar agreement between model prediction and experimental data is observed for the "high-pH" experimental setup (*SI Appendix*, Fig. S2)

The Cost of Extra Chromosomes Increases Linearly with the Total Number of Genes They Contain.

The fitness cost of an aneuploid strain is defined as the reduction of its per-individual offspring. In proliferative conditions, this can be proxied by the growth rate difference with respect to the euploid strain, evaluated in the same environmental conditions. An alternative proxy for fitness is the stationary-phase population size in a given condition. We deduced these proxies from both growth rates (9) and full growth curves (19) collected for yeast aneuploid strains grown in rich media and in the absence of external stress (*Material and Methods*).

In both datasets, we found a statistically significant negative linear correlation between the growth rate of aneuploid strains and the total number of genes carried in the exceeding chromosome. This relation is observed (with different slopes) both in strains with disomic chromosomes compared to a haploid (ploidy 1) genomic background (Fig. 3 *A*, *B*, and *C*) and in strains with trisomic chromosomes compared to a diploid genetic background (Fig. 3*C*). Notably, the same trend is not only observed in aneuploid strains carrying only a single duplicated chromosome (Fig. 3*A*) but also in strains with up to 8 duplicated chromosomes (Fig. 3 *B* and *C*), suggesting that epistatic interactions between the fitness costs of multiple duplicated chromosomes are small. The dataset from ref. 19 also shows a negative correlation between the fitness proxied by stationary-phase population size (optical density, OD) and the total number of genes carried in the excess chromosomes (Fig. 3*B*). Linear negative correlations between growth rates and the number of genes in excess chromosomes of aneuploid strains are also coherently observed in all the stress conditions investigated in the dataset from ref. 9 (*SI Appendix*, Fig.S5A).

Altogether, this experimental body of evidence suggests a general fitness cost for aneuploid individuals with extra chromosomes with respect to a euploid background, of the form

$$\sigma_c = c_0 n_g,\qquad [4]$$

where n_g is the total number of genes carried in the extra chromosomes and c_0 is the average cost per gene, which depends on the external condition and on the background. The fitness described by Eq. (**4**) does not necessarily imply that all the exceeding genes contribute to fitness but also supports a scenario where only a fraction of the genes of the duplicated chromosome can reduce the reproductive fitness (48, 49). The statistically significant linear correlations observed in Fig. 3, however, suggest that the subset of genes that contribute to the cost should be (roughly) evenly distributed across the genome. Indeed, only in such a case, the probability of finding a gene contributing to the fitness cost in a given chromosome would be proportional to its length, giving rise to the observed correlations. We note that this model considers chromosome copy number and duplication of different chromosomes equivalent in terms of per-gene cost. Specifically, the average fitness costs (c_0) of aneuploid strains with diploid vs haploid background display a linear correlation (*SI Appendix*, Fig. S5B), suggesting the existence of a condition-specific effect on the fitness cost. Values of the fitness cost in the diploid background are found to be about a factor one half of those observed in the haploid background, indicating that the development of extra chromosomes is suppressed in haploids and is more likely in diploids, an effect that is in agreement with observations based on evolutionary genomics data (21, 50, 51).

Of note, in Fig. 3*A*, the disomy of Chr VI shows the largest deviation from the linear decreasing trend (similar deviation was observed in ref. 21). This deviation results from an additional fitness cost that is specific to this disomy, which was reported to be lethal in the ploidy = 1 background (19). This additional cost is due to the two key cytoskeleton genes TUB2 (tubulin) and ACT1 (actin), which reduce cell viability when their expression is increased (52, 53). Notably, the effect of the disomy of Chr VI is alleviated in combination with other aneuploidies, for example, Chr I and Chr XIII (19).

A Minimal Fitness Model for Aneuploid Strains with Extra Chromosomes.

The analyses reported in Figs. 2 and 3 suggest that a global effect of extra chromosomes on the growth rate of a strain is recapitulated by a minimal fitness-landscape model, with no epistatic interactions between genes of the extra chromosomes. In this model, fitness is the sum of two contributions: i) a
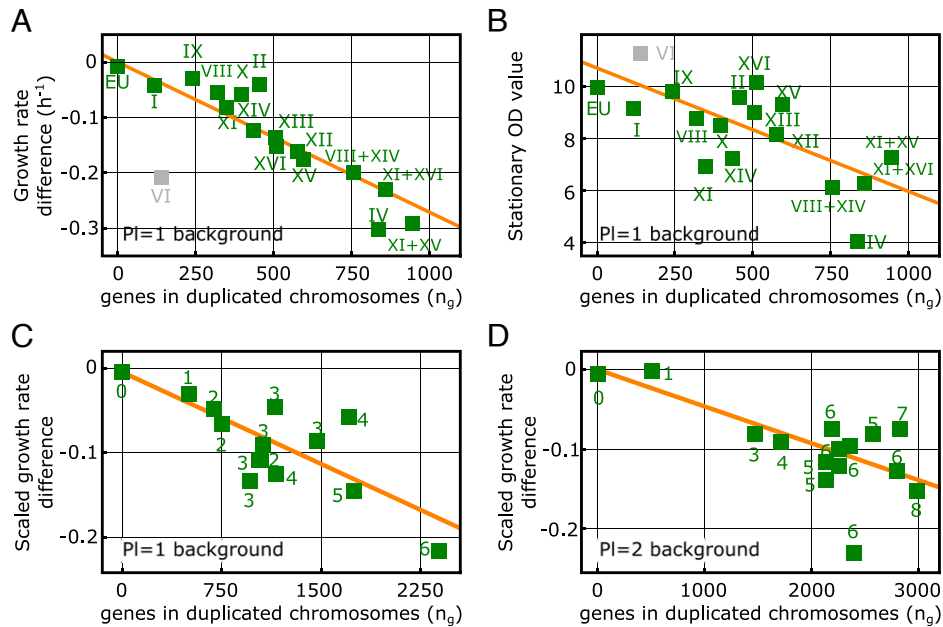
Fig. 3. The fitness cost of extra chromosomes is proportional to the total number of genes present in the excess chromosomes. (*A*) Plot of the values of the difference between the exponential growth rates of aneuploid strains with extra chromosomes and the exponential growth rate of the euploid strain (squares, labels indicating disomic chromosome numbers) against the number of genes carried in the disomic chromosomes (data from ref. 19; *Materials and Methods* for details). The growth rate differences (estimating fitness differences) display a significant negative linear correlation with the number of duplicated genes (red line, Pearsons' r = −0.93, P-value< 10^{-6}). (*B*) Values of the stationary-phase optical density (OD, squares, labels indicate disomic chromosomes) shown against the number of genes in the disomic chromosomes (data from ref. 19). The stationary-phase OD of aneuploid strains, a complementary proxy for the fitness, displays a significant negative linear correlation with the number of duplicated genes (red line, Pearson's r = −0.68, P-value < 0.005). In panels *A* and *B*, the data corresponding to the disomy of Chr VI were not included in the statistical evaluation, as this disomy in a euploid background is known to be lethal on its own (19, 21, 52, 53). (*C* and *D*) Scaled growth rate differences of aneuploid strains obtained from ref. 9 (*Material and Methods*). The plots show scaled growth rate differences (squares) between aneuploid and haploid (*C*) or diploid strains (*D*) against the number of genes carried in unbalanced chromosomes (disomic chromosomes for the left plot and trisomic chromosomes in the right plot). Numbers next to the squares indicate the number of unbalanced chromosomes carried in each strain. In both panels, the proxied fitness difference of aneuploid strains displays a significant negative linear correlation with the number of genes carried in extra chromosomes red lines, Pearson's r = −0.77 (*B*), −0.69 (*C*) and P-value ≤ 0.001 (*B*), 0.004 (*C*).

fitness cost (σ_c) that captures the empirical observation described by Eq. **4**, and ii) a chromosome-specific fitness component, which captures the additional beneficial or detrimental effect of excess chromosomes $\sigma_{\text{kar},s}$. Under these assumptions, the selection coefficient of an aneuploid strain (*s*) in any given growth condition (environment or stress), with respect to the closest euploid background (haploid or diploid) takes the form

$$\sigma_s^{\text{cond}} = -\sigma_{c,s} + \sigma_{\text{kar},s}$$
$$= -c_0 \sum_i \chi_s^i n_i + \sum_i \chi_s^i \sigma_i^{\text{cond}}, \qquad [5]$$

where the karyotype of the strain is defined by the characteristic matrix χ_s, where $\chi_s^i = 1$ if, in the strain *s*, the ith chromosome number exceeds the background ploidy number. The fitness cost of the strain is due to the total number of exceeding chromosomes, $\sigma_s^{\text{cond}} = c_0 \sum_i \chi_s^i n_i = c_0 n_s$, where $c_0 > 0$ is the condition-specific average fitness cost per gene, n_i is the number of genes in the ith chromosome, and n_s is the total number of extra chromosome of strain *s*. Each aneuploid chromosome has an effect on the growth rate σ_i^{cond}, which can either be beneficial ($\sigma_i^{\text{cond}} > 0$) or detrimental ($\sigma_i^{\text{cond}} < 0$) and is condition-specific. This results in the karyotype fitness component $\sigma_{\text{kar},s} = \sum_i \chi_s^i \sigma_i^{\text{cond}}$. A condition (environment or stress) is defined by the value c_0 and the set of values $\{\sigma_i^{\text{cond}}\}$.

The Fitness Landscape Defined by Eq. 5 Captures Nontrivial Behavior of Stress Phenotypes. The two fitness components

of the minimal model, i.e., the fitness costs $\sigma_{c,s}$ and the chromosome-specific fitness effects $\sigma_{\text{kar},s}$, can be inferred from large-scale studies of aneuploid yeast phenotypes in stressful conditions, such as ref. 9 (*Material and Methods*).

The first component captures the global linear decreasing trend of the growth rates of aneuploid strains vs. the total number of exceeding genes, as discussed above. Interestingly, this component can also explain in quantitative terms the observed linear correlation between the degree of phenotypic variation and the degree of overall growth suppression, observed in the data (9) and modeled in refs. 28, 30. In our modeling framework, this correlation corresponds to a linear relationship between the average value ($\overline{\sigma_c}$) and the SD ($\Sigma \equiv \overline{(\sigma_c - (\overline{\sigma_c}))^2}^{\frac{1}{2}}$) of the fitness cost evaluated in a cohort of aneuploid strains. Here, we have denoted with $\overline{\cdot}$ averages computed over the cohort of aneuploid strains in a given growth condition. The fitness cost Eq. **4** predicts a linear relationship of the form

$$\overline{\sigma_c} \sim CV(n_g)\,\Sigma, \qquad [6]$$

where $CV(n_g) = \frac{\text{St.dev}(n_g)}{\overline{n_g}}$ is the coefficient of variation of the distribution of the number of exceeding chromosomes (the total number of genes contained in aneuploid chromosomes) evaluated in the set of aneuploid strains considered. The quantitative expression Eq. **6** explains about 80% of the observed variability of the growth rates in the dataset of ref. 9, implying that the fitness cost alone cannot explain the whole range of observed phenotypic diversity (*SI Appendix*, Fig. S6).

The deviations from this linear trend are captured by the second (condition- and chromosome-specific) fitness component ($\sigma_{\text{kar},s}$), where the effect of an aneuploid extra chromosome (i) on the growth rate is quantified by a chromosome- and condition-specific fitness effect, σ_i^{cond}. *SI Appendix*, Fig. S7A reports the inferred values of the fitness-gain component of each chromosome across stress and control growth conditions for the Pavelka et al. dataset. These inferred values are net of possible confounding factors due to the per-gene fitness cost highlighted previously. Curiously, the chromosome-specific fitness components in the same environment are generally different (uncorrelated) between the ploidy = 1 and ploidy = 2 backgrounds (*SI Appendix*, Fig. S7B). This difference could suggest that the effect of chromosome duplication is ploidy specific, consistently with observations of other ploidy-specific fitness effects (54). However, in the Pavelka *et al.* experiments, each strain was carrying mode than one duplicated chromosome; hence, epistatic interactions between different extra chromosomes could have been present. Unfortunately, the current data are too sparse to infer such epistatic interactions. Additionally, the difference of the chromosome-specific fitness effect between the ploidy = 1 and ploidy = 2 backgrounds could be related to different physiological constraints seen by haploids and diploids and to the different relative gene-dosage increase resulting after a duplication in the two different backgrounds.

Looking at *SI Appendix*, Fig. S7A, one clearly sees that adding different specific extra chromosomes can improve or decrease the fitness in a specific environment, but each environment is characterized by the extent of such fitness gains and losses. For example, in stressful environments such as in the presence of 4NQO, adding an extra chromosome to the genetic background could improve or decrease the fitness by a factor that is more than 10-fold larger than performing the same operation in a nonstressful condition such as glycerol media. Because of this property, it is tempting to classify the "harshness" of an environment by the variability in behavior of aneuploids bearing specific extra chromosomes. Indeed, the variability of effects across chromosomes in a fixed given condition is found to be proportional to the fitness cost per gene observed in the same environment (*SI Appendix*, Fig. S8 *A*, *B*, *C*, and *D*), which can be seen as an independent evaluation of the harshness of that environment. Additionally, contrary to the effects of specific chromosomes, the distributions of the fitness components for the ploidy = 1 and ploidy = 2 backgrounds (shown in *SI Appendix*, Fig. S8 *C*, *D*, and *E*) share common properties that are related to the growth condition. In particular, each environment is characterized by distributions of chromosome-specific fitness effects that have a similar width for ploidy = 1 and ploidy = 2 backgrounds (*SI Appendix*, Fig. S8E).

Expected Interpopulation Dynamics of Aneuploids. The minimal fitness-landscape model described by Eq. **5** can be used to describe the expected interpopulation dynamics of aneuploid strains with a single chromosome gain, by investigating the substitution dynamics associated with Eq. **1** in the landscape (Eq. **5**) when $\chi_s^i = \delta_{i,j}$, for some $j > 0$, and $\delta_{i,j}$ is the standard Kronecker delta. Following a standard population dynamics approach (55, 56), we can use a probabilistic framework to characterize the selective effects of a generic environment on the growth rate of a strain with an aneuploid chromosome, by assuming that the beneficial effect (σ_b) is exponentially distributed, $P(\sigma_b) = be^{-b\sigma_b}$. Averages with respect to this distribution, denoted with $\langle\langle.\rangle\rangle$, quantify the expected dynamics of aneuploid strains with excess chromosomes

in a set of conditions. Hence, they can be used to generate predictions on the typical population dynamics of aneuploids. Under these assumptions, the model predicts that the average probability of developing aneuploidy with extra chromosomes,

$$P_a^{\text{inter}} \equiv \langle\langle \mathcal{P}_a \rangle\rangle \simeq e^{-b\sigma_b^*} = e^{-b\frac{c_0 n_g}{1-r}}, \qquad [7]$$

decreases exponentially with the number of genes contained in the extra chromosomes, suggesting in particular that the relative abundance of duplicated chromosomes is exponentially suppressed with their length. In addition, the typical selection coefficient of an aneuploid strain that has reached fixation,

$$\sigma_b^{\text{inter}} \equiv \frac{\langle\langle \mathcal{P}_a \sigma_b \rangle\rangle}{\langle\langle \mathcal{P}_a \rangle\rangle} \simeq \frac{1}{b} + \sigma_b^* = \frac{1}{b} + \frac{c_0 n_g}{1-r}, \qquad [8]$$

is expected to increase linearly with the cost of extra chromosomes, implying in particular that longer chromosomes require higher fitness advantage to reach fixation.

Combined together, the two model predictions (Eqs. **7** and **8**) suggest an "equilibrium" distribution of aneuploid strains of the form (*Material and Methods*)

$$X_{eq}(n_g) = \frac{1}{Z} \frac{e^{-\kappa n_g}}{n_g}, \qquad [9]$$

where n_g is the number of genes contained in the aneuploid chromosome, Z is a normalization factor, and κ is an effective fitness cost per gene (*Material and Methods*). This prediction is in good agreement with the relative abundances of yeast aneuploid strains observed in evolutionary genomics data (Fig. 4). Interestingly, the numerical values of the effective fitness cost per gene are in agreement with existing experimental evidence (57, 58) suggesting a reduced fitness cost for wild strains (collected as "natural strains" in refs. 57, 59 as "wild strains" in ref. 58). In other words, these strains have a higher propensity to generate aneuploidy, when compared to strains of other kinds, including domesticated, industrial, and human-associated strains (*SI Appendix*, Table S3). We find similar results when comparing the abundance of strains with a ploidy > 2 background to that
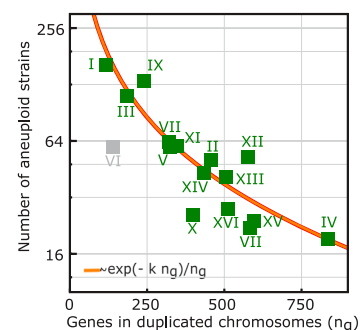
Fig. 4. The fitness landscape derived from phenotyping of laboratory yeast strains explains the relative abundances of yeast aneuploid strains observed in evolutionary genomics data. Relative abundances of aneuploid strains vs. the number of excess genes contained in the aneuploid chromosome (squares) are shown. The numbers of aneuploid strains were retrieved from published data collected in eight studies and reported in ref. 21 (*SI Appendix*, Table S3). The orange line shows the fitness model expectation, for the relative equilibrium frequencies set by chromosome acquisition and loss rates of Eqs. (**7** and **8**), which predicts a functional dependence of the relative frequencies on the number of excess genes (n_g) of the form $\propto \exp(-\kappa n_g)/n_g$. Numerical values of model parameters are reported in *SI Appendix*, Table S3. Data count of the duplications of Chr VI (gray square) was not considered in the model fit since this chromosome is known to be lethal because of the specific effects of the main cytoskeletal genes tubulin and actin (19, 21, 52, 53).

of strains with a lower ploidy background, finding that extra chromosomes are associated with a lower fitness cost in a ploidy > 2 background (*SI Appendix*, Table S3).

Discussion

In yeast, the development of aneuploidy resulting from an accidental chromosome missegregation has been characterized by massive experimental data (3–10, 19–22). As a consequence of this major effort, we are in need of unifying principles to rationalize this wealth of data and embed the underlying evolutionary dynamics into simple quantitative models. Here, we have focused on a specific question, the role of chromosomal duplication with respect to a reference euploid background. Our results show that a simple evolutionary model where a fitness cost of chromosome duplications is counterbalanced by a fitness advantage from the expression of specific genes can explain in quantitative terms two key observations of the emergence dynamics of aneuploidy: i) chromosome duplications emerge transiently as a "quick fix" to dosage insufficiency of a single gene in stressful environments (11, 19, 60) ii) depending on the nature of the applied stress, aneuploidy or local mutations may be favored (20).

While traditionally the fitness advantage of a phenotype associated with a certain mutational target was considered to be the primary trait related to its adaptive value, the recent debate has challenged this assumption based on experimental results that highlight an important role of mutational paths and mutation rates. Our analysis of aneuploids with extra chromosomes provides another example where mutational paths with high rates may give a more relevant contribution to adaptation than mutations with large benefits occurring more rarely (33–35).

Our results support the existence of a cost of single-chromosome duplications that is proportional to the number of genes contained in the exceeding chromosomes. This simple behavior is surprising due to the numerous documented complex physiological changes that emerge with aneuploidies, such as dosage imbalance, effects on interaction networks, and consequent osmotic effects (7, 61). Importantly, our results are in line with a scenario where only a fraction of the genes in a duplicated chromosome will actually contribute to the fitness cost, as supported by the results of refs. 48 and 49. These studies also find that the fitness costs can be complex and specific to a genetic background and that in many cases, they are due to stoichiometric imbalance between proteins that are interaction partners, as supported by recent investigations of the effect of chromosomal imbalance on gene expression, which provided evidence of transmodulations across the genome in aneuploid individuals in both yeast and *Arabidopsis* (62). Of note, such effects were detected only after a careful reanalysis of the yeast transcriptomic dataset collected in ref. 19 and were originally missed because of an unsound normalization of the data (62). More specifically, this scenario implies that some duplicated genes may give a negligible contribution to the fitness cost; hence, the average fitness (c_0) should be thought of as the average of a bimodal distribution (genes with a nonnull contribution plus genes giving a null contribution). In addition, our analysis (Fig. 3) would suggest the class of genes contributing to the cost to be evenly distributed across the genome. This model interpretation is in agreement with existing literature for yeast (48, 49, 63–65). The same interpretation is in accordance with similar effects observed in multiple eukaryotes (66) and could possibly describe other systems, as it is mostly related to general physiological and physical-chemistry principles that should hold across taxa (66). Moreover, the remaining fraction of genes that do not contribute to the cost would be described by our model as behaving neutrally and could therefore be retained in small segmental duplications; this aspect would be in agreement with the conclusions of ref. 63.

Another interesting interpretation of this form of the fitness cost is that the reduction of the growth rate of aneuploid strains may be at least in part the result of the interdependence between growth rate and gene expression, in accordance with the phenomenological laws first observed in bacteria (67), and more recently also in yeast (68). This quantitative framework could also explain the dependence of the fitness cost on the growth conditions that we observe in (Fig. 3). Indeed, the cellular growth rate was shown to be determined by the fraction of proteome occupied by ribosomes, which, in turn, depends on the growth conditions (e.g., nutrient quality). In a similar way, the growth defect due to the overexpression of unneeded proteins was shown to be condition dependent (67, 69). Targeted experiments with combined measurements of growth rate and proteomic allocation in aneuploid strains could be used to test this interpretation of the fitness cost. A possible role of resource allocation in the fitness cost of overexpressed proteins was also suggested in a direct investigation of the cost of overexpressed proteins (49), which, however, also found that these effects vary considerably with the genetic background. Genes contained in an extra chromosome are unnecessary for the survival, and their expression induces a reduction of the growth rate by effectively decreasing the fraction of resources allocated to the ribosomal and housekeeping protein sectors, leading to a decrease in growth rate. This connection holds only if the genetic gene dosage is proportional to gene expression, an effect experimentally observed in yeast (9, 19, 70). Interestingly, the connection between the fitness cost and gene dosage is coherent with our analysis since we observe that values of the fitness cost in the diploid background are close to one half that of those observed in the haploid background, suggesting a connection of the fitness cost per gene to the relative extra gene dosage. Importantly, the linear (per-gene) cost of duplicated chromosomes appears to be a common unifying feature of the fitness landscape in different conditions and environments.

Our model assumes that, when duplicated, some genes will impose a fitness cost to the cell, causing a reduction of its proliferation rate. However, the biological mechanisms that cause this cost are not described within the model, leaving room for several interpretations and further modeling efforts. One contribution to this cost is related to regulatory effects and could be associated with all the genes that are up-regulated together with a duplicated gene, as a result of both cisregulations or transregulations, and additionally, as discussed above, stoichiometric imbalances in protein interaction networks caused by dosage changes are found to play an important role (by regulatory as well as biophysical effects).

For this contribution, the proportionality observed in Fig. 3 would suggest the number of genes affected to be proportional to the number of duplicated and costly genes (i.e., the number of genes that are duplicated and contribute to the cost), meaning that what matters for the main trend is the average number of interaction partners of each such gene.

The formulation of a minimal fitness-landscape model (Eq. **5**) informed by data allows for the inference of chromosome-specific fitness effects. Such effects are related to the dosage increase of genes contained in specific duplicated chromosomes, offering a quantitative framework for the inference of fitness components of early aneuploids. In addition, we have shown that

a simple description of the longer-term evolutionary dynamics of our model (Eqs. **1** and **3**) in this landscape captures the relative abundance of aneuploidies observed in yeast population genomics data. Hence, this model can be used to investigate both the intrapopulation dynamics of aneuploidy individuals within an evolving population (Fig. 2) and the substitution dynamics at the interpopulation population level (Fig. 4).

Importantly, our model can be used to design evolutionary experiments to investigate key biological questions related to the emergence of aneuploidy, which require precise quantitative assessments. For example, our model could be used to test whether the karyotype state of aneuploid individuals is stable in the long term (and in which conditions). As we have shown (*SI Appendix*, Fig. S9), for this question to be addressed, it would be important to design experiments where clonal interference effects would be expected since an observed dynamics with initial rise of aneuploids followed by its elimination from the population could be misinterpreted as a signal of karyotype instability, while simply being the signature of CI. Moreover, our modeling framework could be deployed to design and investigate mutation accumulation (MA) experiments aimed at measuring the missegregation rate. In particular, our quantitative expression for the fitness cost of aneuploid individuals (Fig. 3 and Eq. **4**) could be used to account for fitness effects in MA setups and correct the estimate of the missegregation rate (71, 72).

The model introduced here can be extended to describe more complex scenarios. First, it can be applied to investigate the evolutionary consequences of a sudden increase of the missegregation rate (6), which can result from the usage of antimitotic drugs. Second, it can be used as an ingredient to build models of chromosomal instability, with a clear interest for cancer development (73). For the latter aspect, it would be important to clarify to what extent the increase of gene dosage translates into protein production in human cells (22).

Materials and Methods

Simulations of the Evolutionary Model. We performed numerical simulations of a standard Wright–Fisher model with mutations and selection, with constant population size N. Individuals of the populations are grouped into three distinct and nonoverlapping classes: a) euploid individuals, b) aneuploid individuals, and c) euploid individuals with point mutations. Class (b) is generated from class (a) with a rate μ_a (per individual, per generation), and its members have a selection coefficient $\sigma_b - \sigma_c$. Similarly, individuals of class (c), characterized by a selection coefficient σ_b, are generated from individuals of class (a) with a rate μ_m (per individual, per generation). The simulation is initialized with all individuals assigned to class (a), and it is stopped when either class (b) or class (c) reaches a frequency $x \geq 0.95$. At the end of each simulation, we recorded the successful class either (b) or (c) and the time of appearance (measured in generations) of the first mutant whose descendants took over the whole population (i.e., the emergence time of their last common ancestor t_{min}).

Evolutionary Parameters for the Experimental Data (20). To quantify the fitness cost of the aneuploid strain investigated in ref. 20, we made use of growth curves of the aneuploid and the euploid strains, evaluated at permissive conditions, i.e., without stress. We then inferred the growth rates $f_{eu}^{30°C}$ and $f_{an}^{30°C}$ with an exponential fit of the corresponding growth curves (*SI Appendix*, Fig. S1A). Similarly, for evaluation of the fitness benefit of the aneuploid strain in the presence of a heat stress (39 °C), we inferred the growth rates $f_{eu}^{39°C}$ and $f_{an}^{39°C}$ (*SI Appendix*, Fig. S1B). The two selection coefficients were then computed from the following set of equations:

$$\sigma_c = 1 - \frac{f_{an}^{30°C}}{f_{eu}^{30°C}} \qquad \sigma_b - \sigma_c = \frac{f_{an}^{39°C}}{f_{eu}^{39°C}} - 1. \qquad [10]$$

Selection coefficients of the aneuploid strain for the experiment performed in high pH were computed analogously, using growth curves of the aneuploid and euploid strain evaluated in permissive (*SI Appendix*, Fig. S1C) and stress conditions (*SI Appendix*, Fig. S1D). Numerical values for the inferred growth rates are shown in *SI Appendix*, Table S1.

To estimate the effective population size of the wells used during the evolution experiment, we used the following argument. The experiment in (20) was performed in 96-well plates, with a max volume per well $\simeq 0.4ml$. During the experiment, cell density in liquid cultures was monitored by optical density at 600 nm and reached a maximum value ~ 1 OD_{600}. Since the value $OD_{600} = 1$ corresponds to approximately 10^7 cells per ml (74), we estimated the effective population size to be $10^7 \gtrsim N \gtrsim 10^6$. In Fig. 2 and *SI Appendix*, Fig. S2, we show results obtained with $N = 10^6$, while in *SI Appendix*, Fig. S3, we show equivalent results computed for $N = 10^7$.

The expected cumulative probability for the emergence of aneuploidy with extra chromosomes was computed as using two model ingredients. The first ingredient is the cumulative distribution describing the probability to have a successful aneuploid mutant (i.e., a mutant that eventually will reach fixation) emerging before time t, which reads

$$\mathcal{C}_a(t) \equiv P(t_{emergence} \leq t) = \mathcal{P}_a(1 - e^{-(\lambda_a + \lambda_b)t}), \qquad [11]$$

where $\mathcal{P}_a$ is the aneuploidy fixation probability (Eq. **1**), while $\lambda_m = \mu_m N \phi(\sigma_b, N)$ and $\lambda_a = \mu_a N \phi(\sigma_b - \sigma_c, N)$ are the fixation rates for the euploid and the aneuploid mutants. The fixation probabilities are computed according to Kimura's expression, $\phi(\sigma, N) = (1 - e^{-2\sigma})/(1 - e^{-2\sigma N})$ (40). The second ingredient is the time to fixation of the aneuploid mutant, which reads (*SI Appendix* and ref. 44)

$$t_{fix}^a = \frac{2 \log(2N(\sigma_b - \sigma_c))}{\sigma_b - \sigma_c}. \qquad [12]$$

Finally, the expected cumulative probability for the emergence of aneuploidy reads

$$\mathcal{C}_a^{exp}(t) \equiv \mathcal{C}_a(t - t_{fix}^a). \qquad [13]$$

Growth Curves Data from Ref. 19 and Inference of Growth Rates. In the dataset collected by ref. 19, yeast strains were grown in liquid cultures, and OD_{600} measurements were taken for several time points. Aneuploidy strains were engineered to harbor two specific genes (HIS3 and KAN), integrated in the two copies of the disomic chromosomes (one per copy). The two genes were also integrated in two chromosomes of the euploid strain. Growth curves were evaluated in a medium that is selective for the two genes (–His+G418 medium), therefore preventing the loss of one of the two disomic chromosomes in the aneuploid strains but is otherwise neutral for other traits and does not induce a fitness difference between euploid and aneuploid strains.

Growth rates were then inferred fitting the growth curves to a logistic model

$$Y(t) = \frac{be^{ft}}{1 + a(e^{ft} - 1)}, \qquad [14]$$

where f is the growth rate of the exponential phase, b sets the initial condition $Y(0) = b$, and $\frac{b}{a}$ quantifies the fitness in the stationary phase of the growth (max OD value). We have fitted the data with a parametric Bayesian model $\log(Y(t)) \sim \mathcal{N}((\log(b) + f t - \log(1 + a(e^{ft} - 1)), \sigma_Y)$, choosing priors $a \sim \mathcal{U}(0, 1)$, $b \sim \mathcal{U}(0, 1)$, $f \sim \mathcal{U}(0.1, 1.1)$, and $1/\sigma_Y^2 \sim \Gamma(0.01, 0.01)$, where the symbol $\sim$ stands for "distributed as", and $\mathcal{N}$, $\mathcal{U}$, and Γ stand for normal, uniform, and gamma distribution, respectively. Model fits to the data are shown in *SI Appendix*, Fig. S4, and inferred model parameters are summarized in *SI Appendix*, Table S2.

Dataset from ref. 9 and Evaluation of Growth Rate Differences. Yeast strains in the dataset collected by Pavelka et al. (9) were grown on solid media plates, and growth data were obtained by automated spot detection and intensity measurements. The dataset included 38 fully isogenic aneuploid yeast strains

with distinct karyotypes and genome contents between $1N$ and $3N$ and 3 strains euploid strains (one for each ploidy). In our analysis, we retained strains whose karyotype can be identified as an aneuploid resulting from chromosome gain; hence, we required the total number of genes contained in the excess chromosomes to be about one half of the total number ($\leq$ 3,200 genes). Aneuploidy strains of this form included the majority of the strains of the original dataset (14 strains with a ploidy = 1 background and 15 strains with a ploidy = 2 background). Since our analysis involved stratification of the data according to the closest euploid strains, aneuploid strains with a ploidy = 3 were discarded as too few for a statistical investigation (less than 5 strains).

Growth rate differences were evaluated as follows. The dataset consisted of values of the optical density (OD) of growth assays, evaluated at the same time (t_{max}), of a set of strains with a similar initial number of cells (N_0). Assuming exponential growth (growth assays that reached saturation were excluded from the analysis by the authors), the OD of a specific strain (s), at a given growth condition (c), can be modeled as

$$O_s^c = N_0\, e^{f_s^c t_{max}}, \qquad [15]$$

where f_s^c is the growth rate of strain s in the growth condition c. OD values were then normalized to the value observed for the euploid strain with a ploidy = 1 background and in the same growing condition, obtaining transformed values

$$\hat{O}_s^c \simeq e^{(f_s^c - f_{EU_1}^c)t_{max}}, \qquad [16]$$

which we have log-transformed to get

$$\hat{\Delta}_s^c \equiv \log(\hat{O}_s^c) \simeq (f_s^c - f_{EU_1}^c)t_{max}, \qquad [17]$$

which are scaled (a-dimensional) growth rate differences between a given strain s and the euploid (pl = 1) control strain, evaluated in the growth condition c. Scaled growth differences with respect to the closest strains were then obtained by difference

$$\Delta_s^c = \hat{\Delta}_s^c - \hat{\Delta}_{EU_1}^c = (f_s^c - f_{EU_1}^c)t_{max} \ \ \text{pl} = 1 \ \text{background} \qquad [18]$$

$$\Delta_s^c = \hat{\Delta}_s^c - \hat{\Delta}_{EU_2}^c = (f_s^c - f_{EU_2}^c)t_{max} \ \ \text{pl} = 2 \ \text{background.} \qquad [19]$$

These values were used to evaluate statistics related to the fitness cost component shown in Fig. 2 and *SI Appendix*, Figs. S5 and S6 as well as for the inference of the chromosome fitness components shown in *SI Appendix*, Figs. S7 and S8.

Inference of Fitness Components from Dataset from ref. 9. We consider a minimal fitness model, where the growth rate f_s^c of an aneuploid strain s in the growth condition c reads

$$f_s^c = f_{EU}^c - f_{cost,s}^c + f_{kar,s}^c \qquad [20]$$
$$= f_{EU}^c - c_0 \sum_i \chi_s^i n_i + \sum_i \chi_s^i f_i^c,$$

where f_{EU}^c is the growth rate of the closest euploid strains to s in the same condition. The karyotype of the strain is defined by the matrix χ, where $\chi_s^i = 1$ if, in the strain s, the ith chromosome exceeds the background ploidy number. The fitness cost of the strain is due to the total number of exceeding chromosomes, $f_{cost,s}^c = c_0 \sum_i \chi_s^i n_i = c_0 n_s$, where $c_0 > 0$ is the condition-specific average fitness cost per gene, n_i is the number of genes in the ith chromosome, and n_s is the total number of genes contained in exceeding chromosomes of strain s. Each aneuploid chromosome has an effect on the growth rate f_i^c, which can either be beneficial ($f_i^c > 0$) or detrimental ($f_i^c < 0$) and is condition specific, that results in the karyotype fitness component $f_{kar,s}^c = \sum_i \chi_s^i f_i^c$. In the minimal model Eq. 21, epistatic interactions between chromosomes are not considered.

For each growth condition of the dataset of Pavelka et al, we have inferred the model parameters c_0 and the chromosome fitness effects $\{f_i^c\}$ as follows. It should be noted that, while Eq. 21 requires growth rates (in units $[\text{time}]^{-1}$), the dataset by Pavelka et al. consisted of scaled, a-dimensional values $f_s^c\, t_{max}$,

where t_{max} is a value that is constant for all the strains considered (the time duration of the growth assay). The model Eq. 21 can therefore be inferred with the considered datasets since all the growth rates are scaled by the same value, and the inferred model parameters of Eq. 21 will be expressed in a-dimensional units.

To estimate the value c_0, we performed a linear fit of the data Δ_c^s (Eq. 19, *Material and Methods*) vs n_s, inferring the linear model

$$\Delta_c^{\text{linear},s} = \Delta_c^0 - c_0 n_s. \qquad [21]$$

The value of Δ_c^0 is a correction to the fitness of the euploid strain. Since the data were normalized to the growth of the euploid strain with pl = 1, in the pl = 1 dataset, we imposed $\Delta_c^0 = 0$, while in the pl = 2, the parameter was set free.

Deviations of the data from the linear model were then used to infer the chromosome fitness effects. We first subtracted the linear model contribution, obtaining the detrended data

$$\Delta_c^{\text{detrended},s} \equiv \Delta_c^s - \Delta_c^{\text{linear},s} \qquad [22]$$
$$= \Delta_c^s - \Delta_c^0 + c_0 n_s.$$

The chromosome fitness components are the solution of the linear system

$$\vec{\Delta}_c^{\text{detrended}} = \chi \vec{f^c}. \qquad [23]$$

Since the matrix χ is sparse, the system of equations Eq. 23 cannot be solved exactly . Hence, we use the approximated least square solution of Eq. 23

$$\vec{f^c} = \chi^+ \vec{\Delta}_c, \qquad [24]$$

where χ^+ is the pseudoinverse of χ, the matrix that specifies the karyotypes of the strains considered. Inferred values of the chromosome fitness components are shown in *SI Appendix*, Figs. S7 and S8. *SI Appendix*, Fig. S10 shows a detailed example of the inference procedure discussed in this paragraph.

Fitness-Landscape Prediction for the Relative Abundances of Aneuploid Strains. We computed the equilibrium distribution for the relative abundances of aneuploidy strains as the ratio between the onset rate (r), i.e., at which aneuploid strains reach fixation, and the loss rate (l), i.e., the rate at which an aneuploidy is lost because euploid individuals reach fixation:

$$\chi_{eq} \propto \frac{r}{l}. \qquad [25]$$

In this context, aneuploidy strains are identified by the number of genes that are contained in the duplicated chromosome; hence, the only dependence on the chromosome identity is via its gene content (n_g). The two rates corresponding to the model considered here (Fig. 1) are defined in terms of the model predictions Eqs. (7 and 8) as follows.

The onset rate can be written as the product of P_a^{inter}, the intra population fixation probability (Eq. 7), and an effective rate μ_{stress}, describing the rate at which yeast populations are exposed to stress conditions that can promote the emergence of aneuploidy

$$r(n_g) = \mu_{\text{stress}} P_a^{\text{inter}} \propto e^{-\frac{c_0}{1-r}n_g}. \qquad [26]$$

The loss rate depends on the environmental condition. If the stress condition that promoted the emergence of aneuploidy is no longer in action, then the population will restore the original euploid strain by losing the duplicated chromosome. In this case, the euploid strain, generated with a missegregation rate per individual μ_a^{off}, has a beneficial selection coefficient $\sigma_{eu} = \sigma_c$ (computed with respect to the aneuploidy individual), and the loss rate will equal the substitution rate

$$l^{\text{no stress}}(n_g) = \mu_a^{\text{off}} N\phi(\sigma_{eu}, N) \simeq 2N\mu_a^{\text{off}}\sigma_c \propto c_0 n_g, \qquad [27]$$

while $\phi(\sigma, N) = (1 - e^{-2\sigma})/(1 - e^{-2\sigma N})$ is Kimura's fixation probability (40). If the stress condition persists, then in the long term, the population

will substitute aneuploids with euploid with point mutations, i.e., the second mutational channel considered in our model. In the case the loss of aneuploidy would be attained by the sequential generation of an euploid individual, with a missegregation rate per individual μ_a^{off} and selection coefficient $\sigma_{eu} = \sigma_c - \sigma_b^{\text{intra}} < 0$, which then generates a mutant with a mutation rate individual μ_m and selection coefficient $\sigma_m = \sigma_c > 0$. Note that selection coefficients are now computed with respect to the aneuploidy individual. These two sequential events are known to take place through the so-called "stochastic tunneling" process (75, 76), that makes possible progression through intermediate deleterious alleles without the population ever experiencing the transient decline in fitness that would necessarily occur with sequential fixation. Hence, in the presence of stress, the offset rate is equal to the tunneling rate (75–78)

$$I^{\text{stress}}(n_g) \simeq 2N\mu_a^{\text{off}}\mu_m \frac{\sigma_c}{\sigma_b^{\text{intra}} - \sigma_c} \propto c_0 n_g + \mathcal{O}(c_0^2), \qquad [28]$$

where we used the expression σ_b^{intra} as in Eq. **8**.

While the exact form of the equilibrium distribution will differ if considering persisting/nonpersisting stress conditions after the fixation of aneuploidy, it can be expressed in general terms as a scaling law (vs s the number of genes contained in the aneuploid chromosome n_g) that is valid in both conditions and takes the form

$$X_{eq}(n_g, k) \propto \frac{e^{-k\, n_g}}{n_g}, \qquad [29]$$

where $k \equiv \frac{c_0}{1-r}$ is defined in terms of the condition-specific fitness cost per gene (c_0) and the mutational bias toward the generation of aneuploidy individuals $r = \mu_m/\mu_a$ (Main Text). To account for the variability of the growing conditions, which reflects in the variability of the parameter k, we assume the set of environmental stresses to be described by a uniform distribution for $k \in [0, 2\kappa]$,

where 2κ is an upper bound for k. By averaging Eq. **29** over this distribution, we find

$$X_{eq}(n_g) \propto \frac{e^{-\kappa\, n_g}}{n_g} + \mathcal{O}(\kappa^2), \qquad [30]$$

whose normalized form corresponds to Eq. **9**. The value of the parameter κ is an effective fitness cost per gene, which is proportional to the max value of c_0 of the set of growing conditions considered. We note that the numerical value of c_0 is expected to be lower than the inverse of the typical chromosome size (the longest chromosome in yeast has $n_g \simeq 600$ genes), supporting the approximations taken in Eqs. **28** and **29**. This approximation is also validated a posteriori, by the numerical values obtained in the model fit (*SI Appendix*, Table S3).

Author affiliations: [a]IFOM ETS (Ente del Terzo Settore) - The AIRC (Associazione Italiana per la Ricerca sul Cancro) Institute of Molecular Oncology, Milano 20139, Italy; [b]Dipartimento di Fisica, Università degli Studi di Milano, Milano 20133, Italy; and [c]Istituto Nazionale di Fisica Nucleare (INFN) sezione di Milano, Milano 20133, Italy

1. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
2. L. Sansregret, C. Swanton, The role of aneuploidy in cancer evolution. *Cold Spring Harb. Perspect. Med.* **7**, a028373 (2017).
3. A. M. Taylor *et al.*, Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* **33**, 676–689 (2018).
4. R. D. Kolodner, D. W. Cleveland, C. D. Putnam, Aneuploidy drives a mutator phenotype in cancer. *Science* **333**, 942–943 (2011).
5. R. J. van Vuuren, M. H. Visagie, A. E. Theron, A. M. Joubert, Antimitotic drugs in the treatment of cancer. *Cancer Chemother. Pharmacol.* **76**, 1101–1112 (2015).
6. M. Pavani *et al.*, Epistasis, aneuploidy, and functional mutations underlie evolution of resistance to induced microtubule depolymerization. *EMBO J.*, e108225 (2021).
7. H. J. Tsai *et al.*, Hypo-osmotic-like stress underlies general cellular defects of aneuploidy. *Nature* **570**, 117–121 (2019).
8. A. Terhorst *et al.*, The environmental stress response causes ribosome loss in aneuploid yeast cells. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 17031–17040 (2020).
9. N. Pavelka *et al.*, Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* **468**, 321–325 (2010).
10. J. M. Sheltzer, A. Amon, The aneuploidy paradox: Costs and benefits of an incorrect karyotype. *Trends Genet.* **27**, 446–453 (2011).
11. A. B. Sunshine *et al.*, The fitness consequences of aneuploidy are driven by condition-dependent gene effects. *PLoS Biol.* **13** (2015).
12. D. A. Thompson, M. M. Desai, A. W. Murray, Ploidy controls the success of mutators and nature of mutations during budding yeast evolution. *Curr. Biol.* **16**, 1581–1590 (2006).
13. A. C. Gerstein, Mutational effects depend on ploidy level: All else is not equal. *Biol. Lett.* **9**, 20120614 (2013).
14. K. Krogerus *et al.*, Ploidy influences the functional attributes of de novo lager yeast hybrids. *Appl. Microbiol. Biotechnol.* **100**, 7203–7222 (2016).
15. C. Zeyl, Experimental studies of ploidy evolution in yeast. *FEMS Microbiol. Lett.* **233**, 187–192 (2004).
16. A. C. Gerstein, S. P. Otto, Ploidy and the causes of genomic evolution. *J. Heredity* **100**, 571–581 (2009).
17. J. M. McLaughlan, G. Liti, S. Sharp, A. Maslowska, E. J. Louis, Apparent ploidy effects on silencing are post-transcriptional at HML and telomeres in *Saccharomyces cerevisiae*. *PLoS One* **7**, e39044 (2012).
18. C. W. Bakerlee, A. M. Phillips, A. N. N. Ba, M. M. Desai, Dynamics and variability in the pleiotropic effects of adaptation in laboratory budding yeast populations. *Elife* **10**, e70918 (2021).
19. E. M. Torres *et al.*, Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* **317**, 916–924 (2007).
20. A. W. Yona *et al.*, Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21010–21015 (2012).
21. C. Gilchrist, R. Stelkens, Aneuploidy in yeast: Segregation error or adaptation mechanism? *Yeast* **36**, 525–539 (2019).
22. J. Zhu, H. J. Tsai, M. R. Gordon, R. Li, Cellular stress associated with aneuploidy. *Dev. Cell* **44**, 420–431 (2018).
23. S. Hwang *et al.*, Consequences of aneuploidy in human fibroblasts with trisomy 21. *Proc. Natl. Acad. Sci. U.S.A.* **118** (2021).
24. B. R. Williams *et al.*, Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. *Science* **322**, 703–709 (2008).
25. S. Stingele *et al.*, Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. *Mol. Syst. Biol.* **8**, 608 (2012).
26. T. Davoli *et al.*, Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **155**, 948–962 (2013).
27. N. A. Graham *et al.*, Recurrent patterns of DNA copy number alterations in tumors reflect metabolic selection pressures. *Mol. Syst. Biol.* **13**, 914 (2017).
28. G. Chen *et al.*, Targeting the adaptability of heterogeneous aneuploids. *Cell* **160**, 771–784 (2015).
29. Y. Li *et al.*, Modeling the aneuploidy control of cancer. *BMC Cancer* **10**, 1–9 (2010).
30. A. Kucharavy, B. Rubinstein, J. Zhu, R. Li, Robustness and evolvability of heterogeneous cell populations. *Mol. Biol. Cell* **29**, 1400–1409 (2018).
31. R. Li, J. Zhu, Effects of aneuploidy on cell behaviour and function. *Nat. Rev. Mol. Cell Biol.* 1–16 (2022).
32. R. A. Fisher, *The Genetical Theory of Natural Selection* (The Clarendon Press, 1958).
33. L. Y. Yampolsky, A. Stoltzfus, Bias in the introduction of variation as an orienting factor in evolution. *Evol. Dev.* **3**, 73–83 (2001).
34. E. I. Svensson, D. Berger, The role of mutation bias in adaptive evolution. *Trends Ecol. Evol.* **34**, 422–434 (2019).
35. M. F. Schenk *et al.*, Population size mediates the contribution of high-rate and large-benefit mutations to parallel evolution. *Nat. Ecol. Evol.* **6**, 439–447 (2022).
36. M. S. Rich *et al.*, Comprehensive analysis of the SUL1 promoter of *Saccharomyces cerevisiae*. *Genetics* **203**, 191–202 (2016).
37. L. Keren *et al.*, Massively parallel interrogation of the effects of gene expression levels on fitness. *Cell* **166**, 1282–1294 (2016).
38. F. Duveau, W. Toubiana, P. J. Wittkopp, Fitness effects of Cis-regulatory variants in the saccharomyces cerevisiae TDH3 promoter. *Mol. Biol. Evol.* **34**, 2908–2912 (2017).
39. A. Nourmohammad, T. Held, M. Lässig, Universality and predictability in molecular quantitative genetics. *Curr. Opin. Genet. Dev.* **23**, 684–693 (2013).
40. M. Kimura, On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713 (1962).
41. P. J. Gerrish, R. E. Lenski, The fate of competing beneficial mutations in an asexual population. *Genetica* **102**, 127–144 (1998).
42. M. M. Desai, D. S. Fisher, A. W. Murray, The speed of evolution and maintenance of variation in asexual populations. *Curr. Biol.* **17**, 385–394 (2007).
43. S. C. Park, J. Krug, Clonal interference in large populations. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18135–18140 (2007).
44. S. Schiffels, G. J. Szöllősi, V. Mustonen, M. Lässig, Emergent neutrality in adaptive asexual evolution. *Genetics* **189**, 1361–1375 (2011).
45. K. Jain, J. Krug, S. C. Park, Evolutionary advantage of small populations on complex fitness landscapes. *Evol.: Int. J. Org. Evol.* **65**, 1945–1955 (2011).
46. Y. O. Zhu, M. L. Siegal, D. W. Hall, D. A. Petrov, Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2310–E2318 (2014).

47. G. H. Romano *et al.*, Different sets of QTLs influence fitness variation in yeast. *Mol. Syst. Biol.* **6**, 346 (2010).
48. K. Makanae, R. Kintaka, T. Makino, H. Kitano, H. Moriya, Identification of dosage-sensitive genes in saccharomyces cerevisiae using the genetic tug-of-war method. *Gen. Res.* **23**, 300–311 (2013).
49. D. Robinson, M. Place, J. Hose, A. Jochem, A. P. Gasch, Natural variation in the consequences of gene overexpression and its implications for evolutionary trajectories. *Elife* **10**, e70564 (2021).
50. N. P. Sharp, L. Sandell, C. G. James, S. P. Otto, The genome-wide rate and spectrum of spontaneous mutations differ between haploid and diploid yeast. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E5046–E5055 (2018).
51. K. J. Fisher, S. W. Buskirk, R. C. Vignogna, D. A. Marad, G. I. Lang, Adaptive genome duplication affects patterns of molecular evolution in *Saccharomyces cerevisiae*. *PLoS Genet.* **14**, e1007396 (2018).
52. B. Weinstein, F. Solomon, Phenotypic consequences of tubulin overproduction in *Saccharomyces cerevisiae*: Differences between alpha-tubulin and beta-tubulin. *Mol. Cell. Biol.* **10**, 5295–5304 (1990).
53. H. Liu, J. Krizek, A. Bretscher, Construction of a Gal1-regulated yeast Cdna expression library and its application to the identification of genes whose overexpression causes lethality in yeast. *Genetics* **132**, 665–673 (1992).
54. E. Zörgö *et al.*, Ancient evolutionary trade-offs between yeast ploidy states. *PLoS Genet.* **9**, e1003388 (2013).
55. J. H. Gillespie, Molecular evolution over the mutational landscape. *Evolution* 1116–1129 (1984).
56. H. A. Orr, The distribution of fitness effects among beneficial mutations. *Genetics* **163**, 1519–1526 (2003).
57. J. Muenzner *et al.*, The natural diversity of the yeast proteome reveals chromosome-wide dosage compensation in aneuploids. bioRxiv (2022).
58. J. Hose *et al.*, The genetic basis of aneuploidy tolerance in wild yeast. *Elife* **9**, e52063 (2020).
59. J. Peter *et al.*, Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* **556**, 339–344 (2018).
60. R. Koszul, B. Dujon, G. Fischer, Stability of large segmental duplications in the yeast genome. *Genetics* **172**, 2211–2222 (2006).
61. A. B. Oromendia, S. E. Dodgson, A. Amon, Aneuploidy causes proteotoxic stress in yeast. *Genes Dev.* **26**, 2696–2708 (2012).
62. J. Hou *et al.*, Global impacts of chromosomal imbalance on gene expression in *Arabidopsis* and other taxa. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E11321–E11330 (2018).
63. L. Hakes, J. W. Pinney, S. C. Lovell, S. G. Oliver, D. L. Robertson, All duplicates are not equal: The difference between small-scale and genome duplication. *Gen. Biol.* **8**, 1–13 (2007).
64. A. Wagner, Energy constraints on the evolution of gene expression. *Mol. Biol. Evol.* **22**, 1365–1374 (2005).
65. R. A. Veitia, S. Bottani, J. A. Birchler, Gene dosage effects: Nonlinearities, genetic interactions, and dosage compensation. *Trends Genet.* **29**, 385–393 (2013).
66. J. A. Birchler, R. A. Veitia, One hundred years of gene balance: How stoichiometric issues affect gene expression, genome evolution, and quantitative traits. *Cytoge. Genome Res.* **161**, 529–550 (2021).
67. M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, T. Hwa, Interdependence of cell growth and gene expression: Origins and consequences. *Science* **330**, 1099–1102 (2010).
68. E. Metzl-Raz *et al.*, Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *eLife* **6** (2017).
69. M. Kafri, E. Metzl-Raz, G. Jona, N. Barkai, The cost of protein production. *Cell Rep.* **14**, 22–31 (2016).
70. T. R. Hughes *et al.*, Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
71. A. Mahilkar, N. Raj, S. Kemkar, S. Saini, Selection in a growing colony biases results of mutation accumulation experiments. *Sci. Rep.* **12**, 1–12 (2022).
72. L. M. Wahl, D. Agashe, Selection bias in mutation accumulation. *Evolution* **76**, 528–540 (2022).
73. N. Savage, Bioinformatics: Big data versus the big C. *Nature* **509**, S66–S67 (2014).
74. J. D. Groves, P. Falson, M. Le Maire, M. Tanner, Functional cell surface expression of the anion transport domain of human red cell band 3 (AE1) in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12245–12250 (1996).
75. M. A. Nowak *et al.*, The role of chromosomal instability in tumor initiation. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16226–16231 (2002).
76. Y. Iwasa, F. Michor, M. A. Nowak, Stochastic tunnels in evolutionary dynamics. *Genetics* **166**, 1571–1579 (2004).
77. M. Lynch, A. Abegg, The rate of establishment of complex adaptations. *Mol. Biol. Evol.* **27**, 1404–1414 (2010).
78. M. Lynch, Scaling expectations for the time to establishment of complex adaptations. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 16577–16582 (2010).
79. S. Pompei, M. C. Lagomarsino, Data and codes for "A fitness trade-off explains the early fate of yeast aneuploids with chromosome gains." Mendeley Data. https://data.mendeley.com/datasets/v5w4nvh9vx/1. Deposited 22 March 2023.