Supplementary Information Appendix for

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

Simone Pompei and Marco Cosentino Lagomarsino

Detailed derivation of the analytical results of the evolutionary model

Onset of Aneuploidy. 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 and the euploid mutant respectively, are stochastic variables, with expected values equal to the inverse of the fixation rates ($\tau_{a,m} \equiv \langle t_{a,m} \rangle = 1/\lambda_{a,m}$), and with exponential probability distribution

$$P(t_{a,m}) = \lambda_{a,m} e^{-\lambda_{a,m} t_{a,m}}, \tag{1}$$

The statistics of the fastest emerging mutant can described by the difference of the two times

$$t_{\text{diff}} \equiv t_a - t_m, \tag{2}$$

whose probability density reads

$$P(t_{\text{diff}}) = \begin{cases} \frac{\lambda_a \lambda_m}{\lambda_a + \lambda_m} e^{-\lambda_a t_{\text{diff}}} & t_{\text{diff}} > 0 \\ \frac{\lambda_a \lambda_m}{\lambda_a + \lambda_m} e^{\lambda_m t_{\text{diff}}} & t_{\text{diff}} \leq 0. \end{cases} \tag{3}$$

The problem of computing the probability for aneuploidy to reach fixation is equivalent to computing the probability for the time difference to be negative ($t_{\text{diff}} < 0$). Clonal Interference effects are captured by the extended condition $t_{\text{diff}} + \delta_a^{\text{fix}} < 0$, where δ_a^{fix} is the effective time to fixation of an aneuploid mutant (see next paragraph). This leads to the expression

$$\begin{aligned} \mathcal{P}_a &= P(t_a + \delta^a_{fix} < t_m) && (4) \\ &= P(t_{\text{diff}} < -\delta^a_{fix}) && (5) \\ &= \int_{-\infty}^{-\delta^a_{fix}} P(t_{\text{diff}}) dt_{\text{diff}} && (6) \\ &= \frac{\lambda_a}{\lambda_a + \lambda_m} e^{-\lambda_m \delta^a_{fix}} && (7) \end{aligned}$$

The waiting time until the emergence of the fastest successful mutant can also be investigated through the statistical properties of the minimum of the two waiting times, $t_{\min} \equiv \min(t_a, t_b)$

$$t_{min} \equiv \min(t_m, t_a), \tag{8}$$

whose expected value reads

$$\begin{aligned}
\tau_{min} &\equiv \langle t_{min} \rangle & (9)\\
&= \int_0^\infty dt_a \int_0^\infty dt_m \; t_a \; P(t_a)\, P(t_m) \theta(t_m - (t_a + \delta^a_{fix})) + & (10)\\
&\quad + \int_0^\infty dt_a \int_0^\infty dt_m \; t_m \; P(t_a)\, P(t_m) \theta((t_a + \delta^a_{fix}) - t_m)\\
&= \frac{1}{\lambda_m} \left(1 - (1 + \lambda_m \delta^a_{fix}) \frac{\lambda_a}{\lambda_a + \lambda_m} e^{-\delta^a_{fix} \lambda_m} \right) & (11)
\end{aligned}$$

Effective time to fixation of the aneuploid mutant. In order to compute the effective time where Clonal Interference effects can take place, we follow the same argument presented in [1, 2]. We denote with $x(t)$ the intra-population frequency of the wild-type (not-mutated and euploid) strain, and with $\sigma_a > 0$ the selection coefficient of the aneuploid mutatant , evaluated w.r.t the wild type. The effective population size is N. In the deterministic limit, i.e., neglecting genetic drift effects, the decline in frequency of the wild-type strain is described by the logistic equation

$$x(t) = \frac{x_0 \, e^{-\sigma_a t}}{1 + x_0 \, (e^{-\sigma_a t} - 1)}, \tag{12}$$

where $x_0 \equiv x(t = 0)$. Interference effects can take place during the deterministic dynamics, while genetic drift effects dominate the dynamics for $x(t) \geq x_0 \simeq 1$ and $x(t) \leq 1 - x_f$. Following the argument presented in [2] we use the boundaries where (Laässig boundaries)

$$x_0 = (1 - x_f) = 1 - \frac{1}{2N\sigma_a} \tag{13}$$

. Hence, the time interval during which an interfering mutation can arise is given by

$$\begin{aligned}
t_{fix} &= -\frac{1}{\sigma_a} \log \left(\frac{x_f(1 - x_0)}{x_0(1 - x_f)}. \right) & (14)\\
&= 2 \frac{\log(2N\sigma_a - 1)}{\sigma_a}. & (15)
\end{aligned}$$

Euploid mutants that can emerge from wild type strains with a mutation rate μ_m during this time interval have expected number

$$\mathcal{M}_m = \mu_m N \int_0^{t_{fix}} x(t)\, dt \tag{16}$$

$$= \mu_m N \frac{\log(2N\sigma_a - 1)}{\sigma_a}. \tag{17}$$

Finally, the number of mutations that can interfere with the aneuploid mutant are given by

$$\mathcal{N}_m = \phi(\sigma_m, N)\mathcal{M}_m \tag{18}$$

$$= \mu_m N\, \phi(\sigma_m, N)\, \frac{\log(2N\sigma_a - 1)}{\sigma_a} \tag{19}$$

$$\equiv \lambda_m \delta^a_{fix}, \tag{20}$$

where $\lambda_m = \mu_m N \phi(\sigma_m, N)$ is the fixation rate of the euploid mutant and we have defined the effective time to fixation as

$$\delta^a_{fix} \equiv \frac{\mathcal{N}_m}{\lambda_m} = \frac{\log(2N\sigma_a - 1)}{\sigma_a} \simeq \frac{\log(2N\sigma_a)}{\sigma_a}. \tag{21}$$

Critical value of the beneficial selection coefficient σ_b^* We find the critical value of the beneficial selection coefficient by solving the equation

$$\frac{\lambda_a}{\lambda_a + \lambda_m} e^{-\lambda_m \delta^a_{fix}} \equiv p_0, \tag{22}$$

where $p_0 = \frac{1}{2}$. We solve this equation using the Haldane's Formula [3] for the fixation probability

$$\phi(\sigma, N) = 2\sigma \rightarrow \lambda(\sigma, N, \mu) = 2N\mu\sigma \tag{23}$$

and we consider the following dynamical regimes

(i) *No Clonal Interference.* This regime corresponds to the mathematical condition $\lambda_m \delta^a_{fix} \simeq 0 \rightarrow e^{-\lambda_m \delta^a_{fix}} \simeq 1$, i.e., the expected number of euploid mutants interfering with the fixation dynamics of the aneuploid mutant are negligible. In this regime, Eq.22 reads

$$\frac{\mu_a(\sigma_b^* - \sigma_c)}{\mu_a(\sigma_b^* - \sigma_c) + \mu_m \sigma_b^*} = p_0, \tag{24}$$

and the solution reads

$$\sigma_b^* = \frac{(1-p_0)\sigma_c}{(1-p_0) - \frac{\mu_m}{\mu_a}p_0}, \tag{25}$$

where $r = \frac{\mu_m}{\mu_a}$.

(ii) *In the Clonal Interference regime.* In this regime we find an approximated solution to Eq.22 by considering the first order expansion $e^{-\lambda_m \delta^a_{fix}} \simeq 1 - \lambda_m \delta^a_{fix}$. With this approximation, Eq.22 reads

$$(\sigma_b^* - \sigma_c)(1-p_0)\mu_a = p_0 \frac{\sigma_b^* \mu_m}{1 - 2\mu_m N \sigma_b^* \frac{\log(2N(\sigma_b^* - \sigma_c))}{\sigma_b^* - \sigma_c}} \tag{26}$$

by approximating $\frac{\log(2N(\sigma_b^* - \sigma_c))}{\sigma_b^* - \sigma_c} \simeq \frac{\log(2N)}{\sigma_b^* - \sigma_c}$ we get the expression

$$(\sigma_b^* - \sigma_c)(1-p_0)\mu_a = p_0 \frac{\sigma_b^* \mu_m}{1 - 2\mu_m N \sigma_b^* \frac{\log(2N)}{\sigma_b^* - \sigma_c}} \tag{27}$$

with the explicit analytic solution

$$\begin{aligned} \sigma_b^* &= \frac{(1-p_0)\,\sigma_c}{(1-p_0) - \frac{\mu_m}{\mu_a}p_0} \left(\frac{1}{1 - 2\mu_m N \log(2N) \frac{(1-p_0)}{1 - p_0\left(1 + \frac{\mu_m}{\mu_a}\right)}} \right) \\ &\simeq \frac{(1-p_0)\sigma_c}{(1-p_0) - \frac{\mu_m}{\mu_a}p_0} + \mathcal{O}(\mu_m N \log(2N)) \end{aligned} \tag{28}$$

Hence, the general solution that recapitulates both dynamical regime is obtained for $p_0 = 1/2$ and reads

$$\sigma_b^* = \frac{\sigma_c}{1-r} + \mathcal{O}(\mu_m N \log(N)) \tag{29}$$

where $r = \mu_m/\mu_a$.

High Ph experiment data from ref. [4]

In ref. [4], Yona and coworkers investigated an evolved strain obtained from an experiment presented in a previous publication [5], in which a haploid S. cerevisiae was evolved under high PH (8.6), using a transfer protocol similar to the one used for the High T experiment. In such experiment (one replicate only) the duplication of chromosome V (trisomy) was

found to have reached fixation after about $\sim$ 150 generations; the duplication of this chromosome shown to confer a beneficial effect in response to the applied stress.

Growth curves relative to this strains (aneuploid and euploid) were obtained from spores produced at the end of the experiments, and were used here to infer growth rates. It should be noted that these two strains, namely the euploid reference and aneuploid strains, contained 4 point mutations that emerged during the experiment and were not present in the euploid strain used at the beginning of the experiment. This set of mutations could possibly confer additional advantage to the high PH [5].

In this context we neglected the effect of these mutations, and we showed that the evolutionary dynamics leading to the emergence of resistance to High-PH can be explained by the emerging dynamics of aneuploidy alone. It should be noted, in particular, that the difference in fitness between the initial euploid strain and the aneuploid strain is higher than the one observed (euploid vs aneuploid, both of them with the set of adaptive mutations). Our estimate of $\sigma_b - \sigma_c$ is therefore a conservative one, since it quantifies the minimal fitness advantage conferred by the chromosome gain alone.

Relaxing the assumption for the selection coefficient of euploid mutant

This section reviews the model assumption used throughout the paper that concerns the selection coefficient of the euploid mutant. We show how our model maps to more complex scenarios within the same model formalism. More specifically, we have assumed that, in response to external stress, an euploid individual developing point mutations can gain a fitness benefit equal to the fitness gain attained by anueuploid individuals

$$\sigma_m = \sigma_b. \tag{30}$$

Using this assumption we have considered a conservative scenario, i.e., the one where the emergence of aneuploidy is mostly suppressed, and where 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)

However, in a more general scenario, one could focus on point mutations that alter gene expression to a different degree (highter or lower). In this case, as described in the main text, the two mutants (anueploid vs euploid with point mutations) would have a different selection coefficient, a situation that can be generically described by the condition

$$\sigma_m = \epsilon\ \sigma_b, \tag{31}$$

where the condition $0 < \epsilon < 1$ describes the case where the fitness gain of the mutant is lower than that of aneuploid, while with $\epsilon > 1$ the fitness gain of the anuploid is lower than the one of the mutant.

With this model extension, the fixation probability of the euploid mutant reads

$$\begin{aligned} \lambda_m &= \mu_m N \phi_m(\epsilon\sigma_b, N) && (32) \\ &= \mu_m^{\mathrm{mod}} N \phi_m(\sigma_b, N), && (33) \end{aligned}$$

where $\phi_m(\sigma_m, N)$ is the fixation probability of the euploid mutant in a population with N individuals and we have definied the modified mutation rate

$$\mu_m^{\mathrm{mod}} \equiv \frac{\phi_m(\epsilon\sigma_b, N)}{\phi_m(\sigma_b, N)} \mu_m \tag{34}$$

and fullfills the condition

$$\begin{aligned} \mu_m^{\mathrm{mod}} &< \mu_m \quad \text{for} \quad \epsilon < 1 \\ \mu_m^{\mathrm{mod}} &> \mu_m \quad \text{for} \quad \epsilon > 1 \end{aligned} \tag{35}$$

In other words, from Eq.s (33,34), one can deduce that a model with a selection coefficient $\sigma_m = \epsilon\sigma_b$ is equivalent to a model with $\sigma_m = \sigma_b$ and a modified mutation rate. Hence, relaxing the assumption Eq.30 corresponds to considering an effective mutation rate μ_m^{mod} which could be different than the per-base spontaneous mutation rate. Of note, when considering the Haldane's Formula [3] for the fixation probability, one has $\mu_m^{\mathrm{mod}} = \epsilon\mu_m$. Considering the comparison of our model with data from the evolutionary experiment [4], we find that the model prediction is in agreement with the experimental data also for values of mutational rates lower than the spontaneous per-nucleotide error rate $\mu_m \leq \mu_{\mathrm{spont.}} = 1.7 * 10^{-10}\mathrm{gen}^{-1}$ (cfr. Fig.2, S2, S3). Hence, our model would also support a scenario where the euploid mutant is characterized by a selection coefficient $\sigma_m \leq \sigma_b$.

Loss of aneuploidy in the clonal interference regime

This section addresses the quantitative predictions of our model that are key for the interpretation of evolutionary experiments aimed at investigating the fate of aneuploid individuals, i.e., testing whether or not this karyotype state is genetically stable. In experimental setups akin to that used in ref. [4], the loss of aneuploidy in the long term could be explained by two alternative scenarios: (i) loss of the aneuploidy and subsequent gain of the point mutations or (ii) elimination of aneuploid individuals resulting from clonal

interference effects (the euploid point mutation emerges before the complete fixation of aneuploidy, and outcompetes it). Scenario (i) would imply that aneuploid individuals are not stable (the extra chromosome is lost) while scenario (ii) does not exclude this possibility (in this case aneuploid individuals did not lose the extra chromosome). In the following, we will give quantitative conditions for scenario (ii) to be observed in experimental setups akin to ref. [4].

To illustrate this point, we consider a conservative scenario where the emergence of aneuploidy would be very likely if no Clonal Inferference (C.I.) effects were present, and corresponds to the mathematical condition

$$\lambda_a \gg \lambda_m \rightarrow \frac{\lambda_a}{\lambda_a + \lambda_m} \simeq 1. \tag{36}$$

This model condition is fulfilled by the model parameters estimated for the experiment of ref. [4] (for which $\sigma_b = 0.17\text{gen}^{-1}$ and $\sigma_b = 0.12\text{gen}^{-1}$ and using realistic values of the mutational rates $\mu_a \simeq 10^{-6}\text{gen}^{-1}$ and $\mu_m \leq 10^{-8}\text{gen}^{-1}$). Hence, this regime describes a realistic scenario for evolutionary experiments akin to [4].

In this case, the probability to develop aneuploidy reads

$$\mathcal{P}_a \simeq e^{-\lambda_m(\sigma_m,\mu_m,N)\delta^a_{fix}(\sigma_a,N)}, \tag{37}$$

and predicts C.I. effects to dominate the dynamics and substantially reduce the probability to develop aneuploidy for experiments carried with an effective population sizes larger than a critical value

$$N^* = \frac{\log[2]}{2\mu_m}\frac{\sigma_a}{\sigma_m}\frac{1}{\mathcal{W}(\frac{\log[2]\sigma_a^2}{\mu_m\sigma_m})}, \tag{38}$$

where $\mathcal{W}(x)$ is the Lambert-W function (note that Eq.(38) is the solution to the equation $e^{-\lambda_m(\sigma_m,N,\mu_m)\delta^a_{fix}(\sigma_a,N)} = 0.5$).

The dynamics predicted by our model in this regime for a typical evolutionary experiment (akin to ref. [4]) is shown in Fig. S9 A. When the effective population size used during the experiment is close to the critical value ($N \simeq N^*$) we find that aneuploids can reach a high intra-population frequency before being replaced by the euploid point mutation (Fig. S9 B.). Hence, in this regime, the prediction of our model is that the loss of aneuploid individuals from the population could be mistaken as a sign of genomic instability while, instead, it is an indirect effect caused by C.I.

References

[1] P. J. Gerrish and R. E. Lenski, "The fate of competing beneficial mutations in an asexual population," *Genetica*, vol. 102, pp. 127–144, 1998.

[2] S. Schiffels, G. J. Szöllősi, V. Mustonen, and M. Lässig, "Emergent neutrality in adaptive asexual evolution," *Genetics*, vol. 189, no. 4, pp. 1361–1375, 2011.

[3] J. B. S. Haldane, "A mathematical theory of natural and artificial selection, part v: selection and mutation," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 23, pp. 838–844, Cambridge University Press, 1927.

[4] A. H. Yona, Y. S. Manor, R. H. Herbst, G. H. Romano, A. Mitchell, M. Kupiec, Y. Pilpel, and O. Dahan, "Chromosomal duplication is a transient evolutionary solution to stress," *Proceedings of the National Academy of Sciences*, vol. 109, no. 51, pp. 21010–21015, 2012.

[5] G. H. Romano, Y. Gurvich, O. Lavi, I. Ulitsky, R. Shamir, and M. Kupiec, "Different sets of qtls influence fitness variation in yeast," *Molecular systems biology*, vol. 6, no. 1, p. 346, 2010.

[6] A. H. Yona, Y. S. Manor, R. H. Herbst, G. H. Romano, A. Mitchell, M. Kupiec, Y. Pilpel, and O. Dahan, "Chromosomal duplication is a transient evolutionary solution to stress," *Proceedings of the National Academy of Sciences*, vol. 109, no. 51, pp. 21010–21015, 2012.

[7] E. M. Torres, T. Sokolsky, C. M. Tucker, L. Y. Chan, M. Boselli, M. J. Dunham, and A. Amon, "Effects of aneuploidy on cellular physiology and cell division in haploid yeast," *science*, vol. 317, no. 5840, pp. 916–924, 2007.

[8] N. Pavelka, G. Rancati, J. Zhu, W. D. Bradford, A. Saraf, L. Florens, B. W. Sanderson, G. L. Hattem, and R. Li, "Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast," *Nature*, vol. 468, no. 7321, pp. 321–325, 2010.

[9] G. Chen, W. A. Mulla, A. Kucharavy, H.-J. Tsai, B. Rubinstein, J. Conkright, S. McCroskey, W. D. Bradford, L. Weems, J. S. Haug, *et al.*, "Targeting the adaptability of heterogeneous aneuploids," *Cell*, vol. 160, no. 4, pp. 771–784, 2015.

[10] C. Gilchrist and R. Stelkens, "Aneuploidy in yeast: Segregation error or adaptation mechanism?," *Yeast*, vol. 36, no. 9, pp. 525–539, 2019.

[11] J. Peter, M. De Chiara, A. Friedrich, J.-X. Yue, D. Pflieger, A. Bergström, A. Sigwalt, B. Barre, K. Freel, A. Llored, *et al.*, "Genome evolution across 1,011 saccharomyces cerevisiae isolates," *Nature*, vol. 556, no. 7701, pp. 339–344, 2018.

[12] B. Gallone, J. Steensels, T. Prahl, L. Soriaga, V. Saels, B. Herrera-Malaver, A. Merlevede, M. Roncoroni, K. Voordeckers, L. Miraglia, *et al.*, "Domestication and divergence of saccharomyces cerevisiae beer yeasts," *Cell*, vol. 166, no. 6, pp. 1397–1410, 2016.

[13] S.-F. Duan, P.-J. Han, Q.-M. Wang, W.-Q. Liu, J.-Y. Shi, K. Li, X.-L. Zhang, and F.-Y. Bai, "The origin and adaptive evolution of domesticated populations of yeast from far east asia," *Nature communications*, vol. 9, no. 1, pp. 1–13, 2018.

[14] Y. O. Zhu, G. Sherlock, and D. A. Petrov, "Whole genome analysis of 132 clinical saccharomyces cerevisiae strains reveals extensive ploidy variation," *G3: Genes, Genomes, Genetics*, vol. 6, no. 8, pp. 2421–2434, 2016.

[15] K. C. Kao, K. Schwartz, and G. Sherlock, "A genome-wide analysis reveals no nuclear dobzhansky-muller pairs of determinants of speciation between s. cerevisiae and s. paradoxus, but suggests more complex incompatibilities," *PLoS genetics*, vol. 6, no. 7, p. e1001038, 2010.

[16] A. M. Selmecki, Y. E. Maruvka, P. A. Richmond, M. Guillet, N. Shoresh, A. L. Sorenson, S. De, R. Kishony, F. Michor, R. Dowell, *et al.*, "Polyploidy can drive rapid adaptation in yeast," *Nature*, vol. 519, no. 7543, pp. 349–352, 2015.

[17] M. Jaffe, G. Sherlock, and S. F. Levy, "iseq: a new double-barcode method for detecting dynamic genetic interactions in yeast," *G3: Genes, Genomes, Genetics*, vol. 7, no. 1, pp. 143–153, 2017.

[18] N. P. Sharp, L. Sandell, C. G. James, and S. P. Otto, "The genome-wide rate and spectrum of spontaneous mutations differ between haploid and diploid yeast," *Proceedings of the National Academy of Sciences*, vol. 115, no. 22, pp. E5046–E5055, 2018.

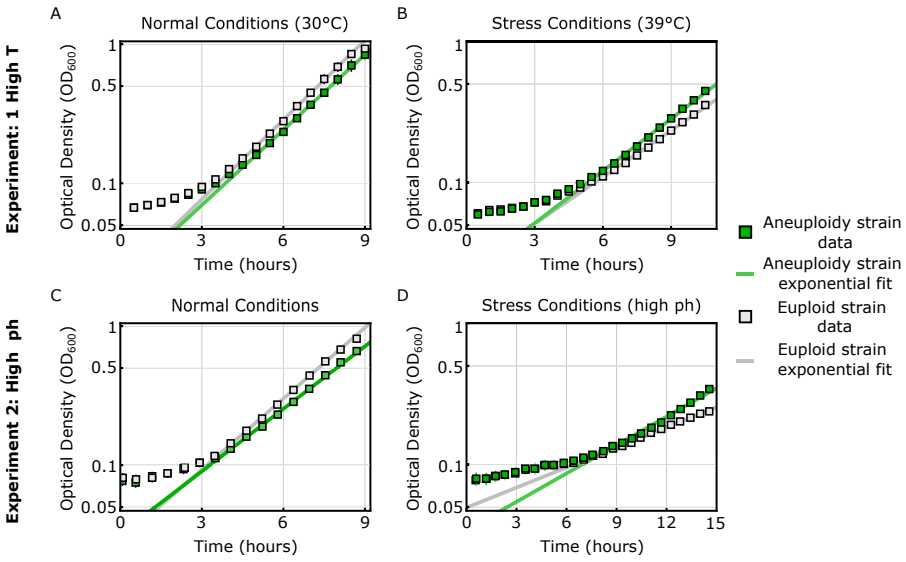

Figure S1: **Growth rates describing of the aneuploid and euploid strain from ref. [6] are inferred from growth curves.** Shown here are growth curves (Optical Density vs time) of the aneuploid strain (diploid strain with the trisomy of chromosome III) and of the diploid strain with the same genetic background evaluated both in normal conditions (30°C, A) and in stress conditions (39 °C, B). Similarly, in C and D we show the growth curves of the aneuploid strain aneuploid strain (diploid strain with the trisomy of chromosome IV) and of the diploid strain with the same genetic background evaluated both in normal conditions (C) and in stress conditions (high ph, D) The experimental data (squares + bars, showing mean and standard deviation evaluated over a set of ~ 35 replicates) is shown together with an exponential fit (solid lines), evaluated neglecting the lag phase (~ 4 hours in A, ~ 5.5 hours in B ,~ 4 hours in C and ~ 7.5 hours in D). Values of the inferred growth rates are reported in Table S1.

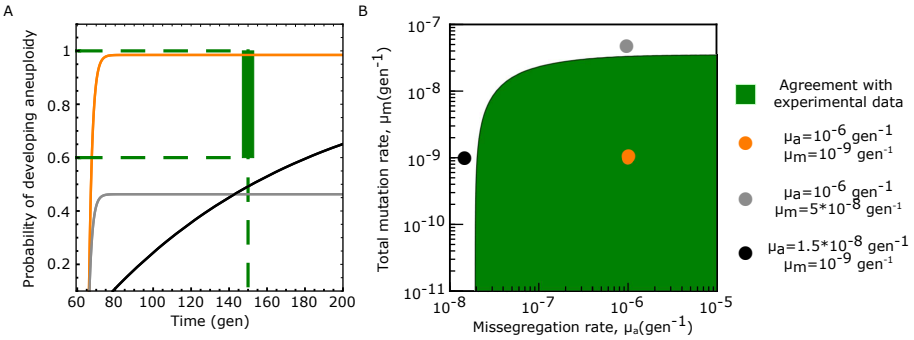

Figure S2: **Model predictions agree with laboratory-evolution data from ref. [6] (High ph experimental setup)**. (A) Expected cumulative probability for the emergence of aneuploidy with extra chromosomes vs the time to reach fixation (see Material and Methods), computed according to the model prediction (Eqs 1, 3, Main text) 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, 1 out of 1 yeast population developed chromosomal duplications ($CI_{66\%} = [0.6, 1]$ for the probability to develop aneuploidy), and the fixation wes reached before 150 generations. Hence, the experimental data fall in region of the plot corresponding to $P_a \in [0.6, 1]$ and $t = 150gen$, marked in green, delineate. Trajectories predicted by the model that cross this region are in agreement with the experimental data. Similarly, in (B) we show the combinations of the numerical values of the model parameters (μ_a, μ_m) that are in agreement with the experimental data, while coloured dots marks the values corresponding to the trajectories shown in A. Numerical values of the beneficial selection coefficient ($\sigma_b = 0.29\,\mathrm{gen}^{-1}$) and for the fitness cost of aneuploidy ($\sigma_c = 0.12\,\mathrm{gen}^{-1}$) were obtained from exponential fits of the growth curves of the corresponding yeast strains [6], (see Material and Methods and Fig. S1). The effective population size was set to $N = 10^6$ individuals (Fig. S3 shows results for $N = 10^7$).

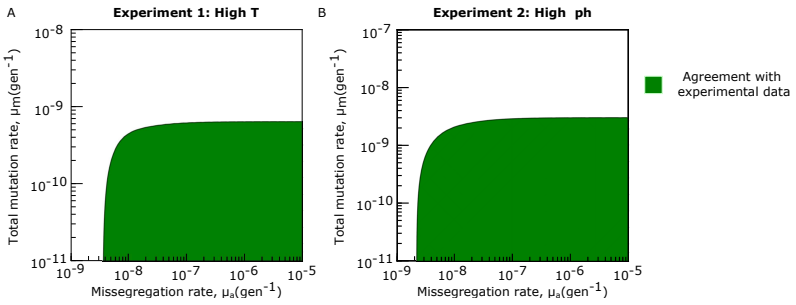

Figure S3: **Quantitative model predictions agree with the laboratory evolutionary experiment from ref. [6] for** $N = 10^7$. Values of the model parameters (μ_a, μ_b) for which the model prediction is in agreement with the experimental data of ref [6] when assuming a population size of $N = 10^7$ individuals, for the experimental setup at high temperature (A) and high ph (B) of agreement b as Fig. 2B

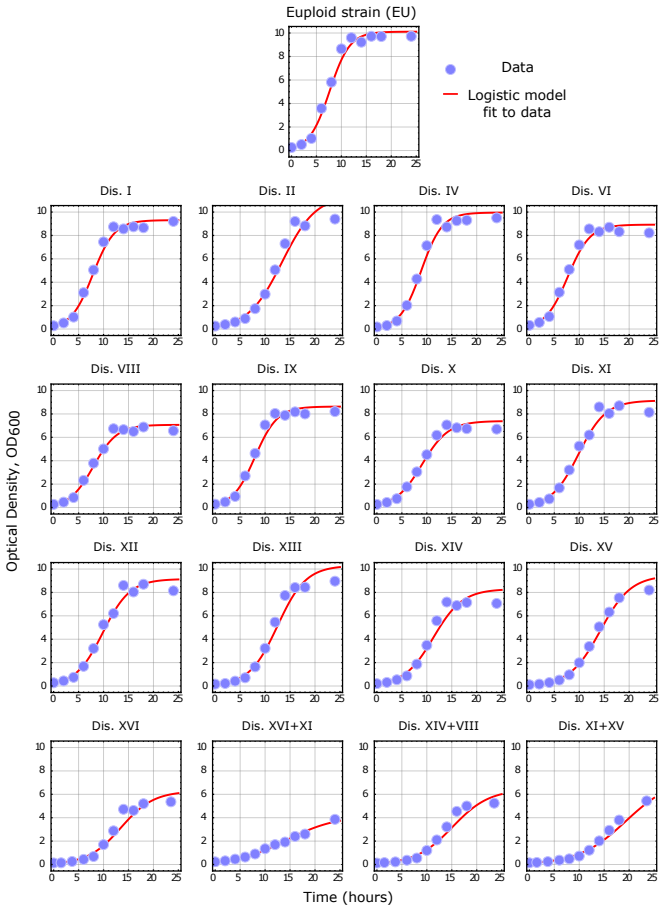

Figure S4: **Growth rates of aneuploid strains from ref. [7] are inferred with a logistic fit of growth curves.** Shown here are growth curves (Optical Density vs time) of the aneuploid strain and of the euploid strains with the same genetic background taken form Torres at al. The experimental data (circles) are shown together with a logistic fit (solid red lines). Model parameters were inferred with a Bayesian framework (see Material and Methods). Values of the posterior mean values are reported in Table S2.

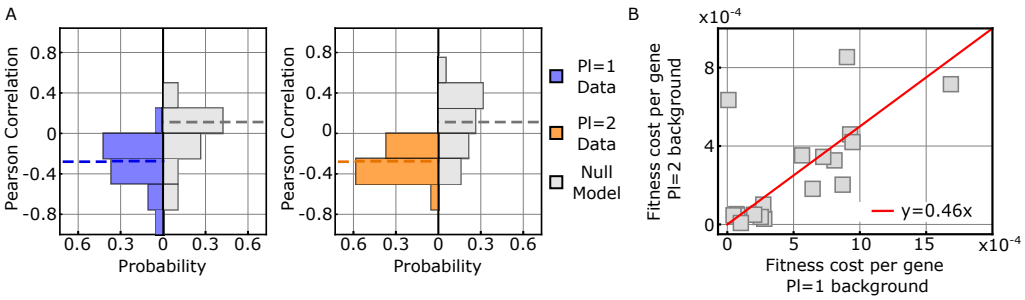

Figure S5: **Linear negative correlations between growth rates and number of genes in exceeding chromosomes of aneuploid strains are coherently observed in all growth conditions.**A: Histograms of the values of the Pearson's correlation coefficients of aneuploidy strains with a Ploidy=1 (Left) and a Ploidy=2 background (Right, for the data-set from ref. [8]. In each condition, the Pearson coefficient was evaluated between the values of strain growth rates and the corresponding total number of genes contained in the aneuploidy chromosomes. The null model was obtained with a randomization test, where, in each conditions, growth rates and number of exceeding genes were randomly shuffled and the Pearson correlation coefficient was evaluated for the shuffled data. Dashed lines mark the mean value of each histogram. The difference between the null model distribution and the corresponding distributions of Pl=1 and Pl=2 background are statistically significant ($p_{val} = 0.00003, 0.005$ respectively, Mann Whitney U test). B: Scatter plot for the fitness cost per gene (c_0) evaluated for Pl=1 and Pl=2 strains (each data-point correspond to the same condition). The values of the fitness cost display a statistically significant linear correlation (Pearson Correaltion coefficient 0.68, $p_{val} < 0.002$). The red line show the best linear fit of the data-points.

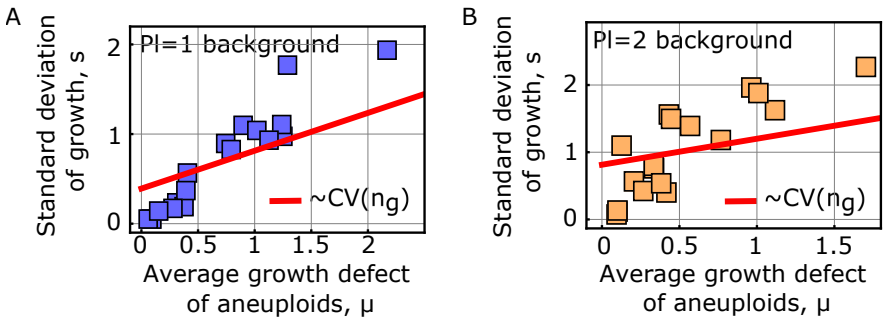

Figure S6: **The fitness cost model predicts a linear relationship between the average fitness defect of aneuploids and standard deviation of growth rates, and explains about** 80% **of the observed dispersion.** A and B: Model predictions capture large-scale phenotype data from ref. [9]. The scatter plots of the average growth rate defect (μ) of an aneuploid strain across several conditions (environments and stresses) versus its standard deviation across the same set of conditions (s). Each square represents a strain of the aneuploid collection of ref. [8] (see Materials and Methods for description of the data-set). Panel A refers to a haploid background, whereas panel B refers to a diploid background. Our form of the fitness cost predicts a linear relation between μ and s, with a slope equal to the coefficient of variation (CV) of the distribution of the number of excess genes (n_g) contained in the aneuploid chromosomes across the set of strains (see Materials and Methods). The red lines show that the this model components alone can describe the observed linear trend and explain the data only partially, with values of the R^2 statistics: $0.84, 0.80$ for panel A and B respectively.

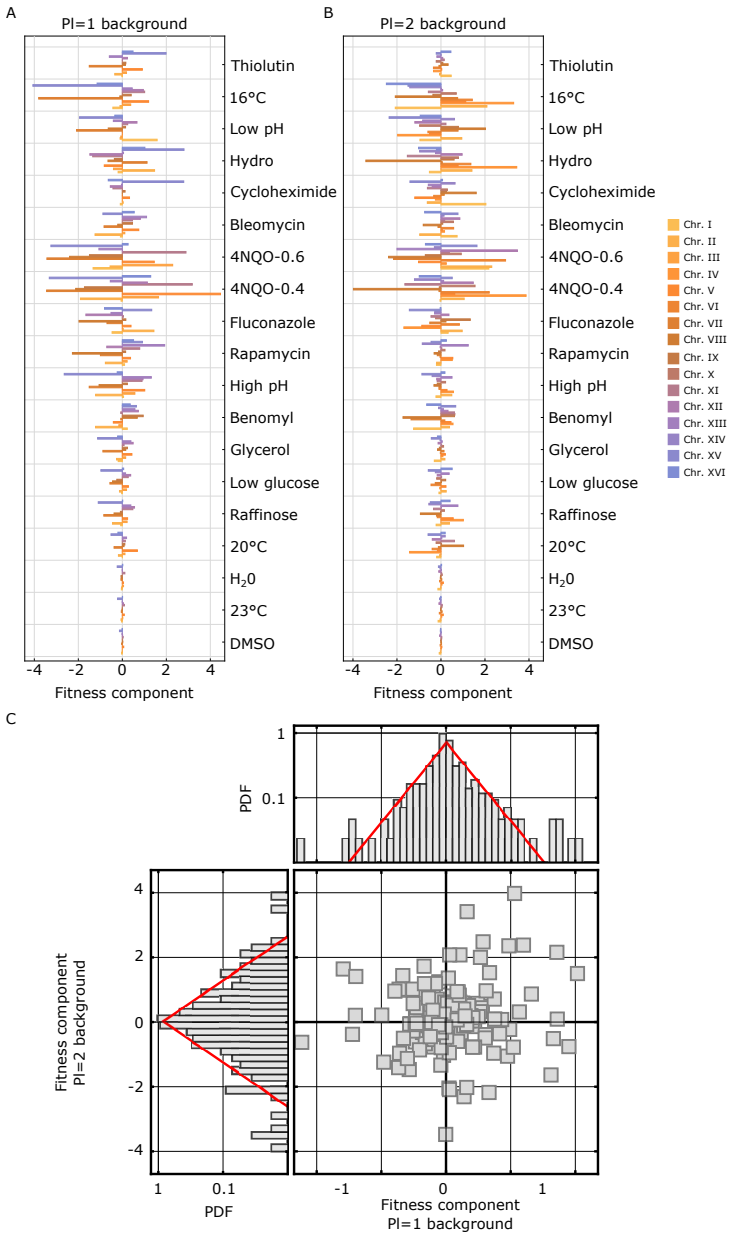

Figure S7: **Chromosome fitness components of aneuploid strains can be inferred from measured growth rates.** The inferred fitness component of A, a disomic chromosome in a ploidy=1 background, and B, a trisomic chromosome in a ploidy=2 background, is shwon here as a function of the stress condition in which growth rates were evaluated (shown in the x axis). In both the two panels, the chromosome-specific component is shown with a bar plot color coded according to the legend shown on the right. Chromosome fitness components were inferred from growth rates obtained from [8]. Details about the inference are given in Materials and Methods.B. Statistics of the inferred values of the selection coefficients of aneuploidy chromosomes, inferred from [8],and evaluated across several stress conditions. The central panel shows the scatter plot for the values of the chromosomal selection coefficient evaluated in strains with ploidy=1 background vs ploidy=2 background (each square correspond to the same condition and the same chromosome). The data does not show statstically significant correlation (Pearson's r=0.11). Left and top panels show the probability distribution density of the chromosomal selection coefficients across all conditions and for all chromosomes together for (Top) ploidy=1 and (Left) ploidy=2 background, respectively (gray boxes). The two distributions are in very good agreement with a Laplace distribution (red lines, mean values $-0.01, 0.02$ and variance $1, 0.7$ respectively). Note that in the central panel we report only components that are present in both data-sets.

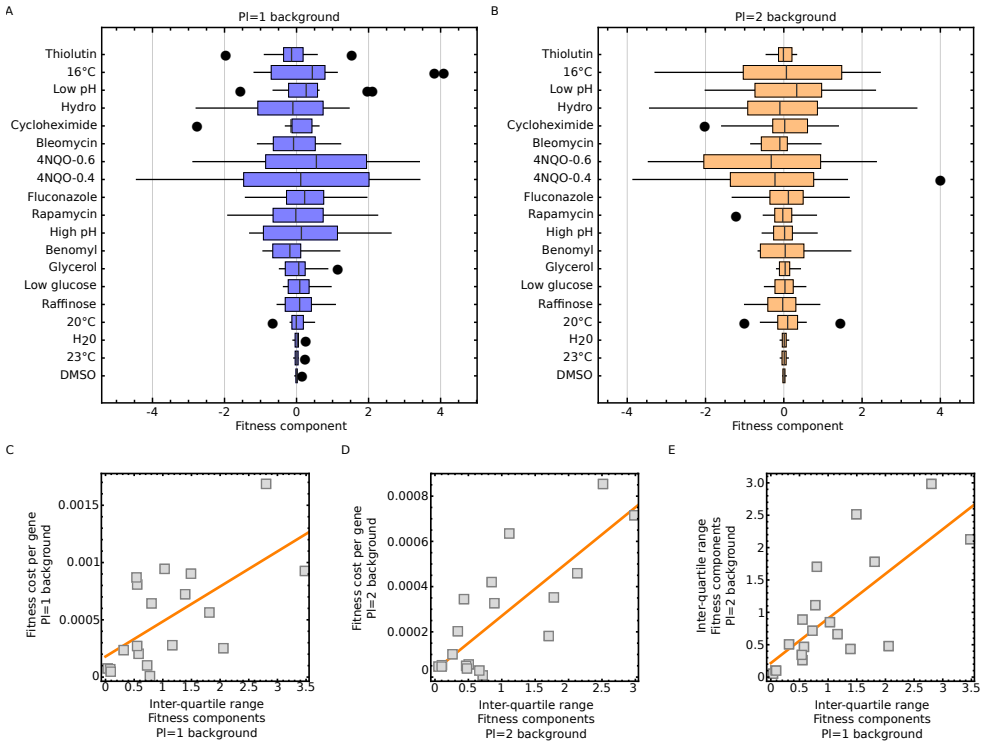

Figure S8: **The statistical properties of the distributions of chromosomal fitness components classify environments by harshness.** Panels A and B show Box-Whisker plots for the distributions of the inferred values of the chromosome-specific fitness components for each environment. Panel A shows the distributions for ploidy=1 background and panel B shows the distributions for ploidy=2 background. In the boxes, the black line marks the mean value of the distribution, the boxes make the lower (Q1) and upper (Q3) quartile. Whiskers mark the minumum-maximum values and circles mark outliers (points beyond the inter-quartile range from the edge of the box). C: Scatter plot of the inter-quartile ranges of the distributions shown in A *vs* fitness cost per gene in the ploidy=1 background, supporting the idea that the environment-specific inter-chromosome variability of fitness effects is a proxy of environmental harshness. D: Scatter plot for the inter-quartile ranges of the distributions shown in B *vs* fitness cost per gene in the pl=2 background. E: Scatter plot between the inter-quartile ranges of the distributions for ploidy=1 and polidy=2 backgrounds, showing that the width of the distributions are correlated. In panels C-D-E we observe a statistically significant linear correlation coefficient, with Pearson's r values= $(0.8, 0.7, 0.6)$ and p-val= $(0.00004, 0.0003, 0.003)$ respectively.

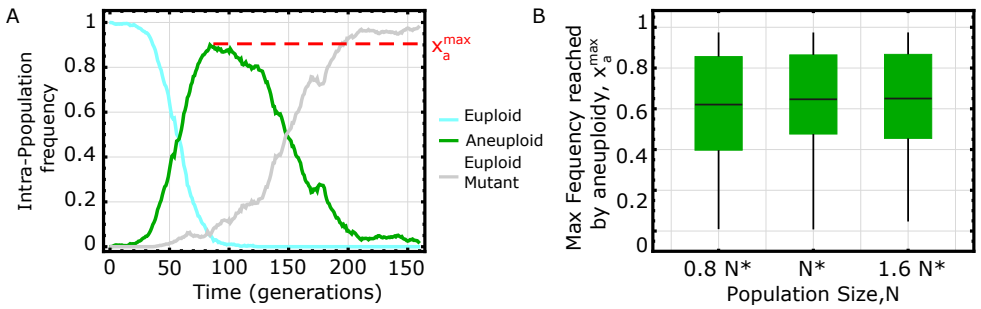

Figure S9: **Clonal Interference (C.I.) effects can be mistaken for a loss of aneuploidy resulting from karyotype instability.** We show here the results of numerical simulations of our evolutionary model (cfr. Material and Methods for details about the simulations)to illustrate the model predictions in the CI regime. In (A) we show one simulation instance, to illustrate the typical dynamics of the intra-population frequencies of euploid individuals (cyan), aneuploid individuals (green) and of euploid mutants (gray). The dynamics displays two distinct phases: (i) a rise in frequency of aneuploid individuals up to a maximum value $x_a^{\max}$, followed by (ii) a decline and subsequent elimination of aneuploidy from the population, because of the emergence of an euploid mutant. B. Probability distribution of the maximal frequency reached by aneyploidy before being eliminated from the population because of CI *vs* the value of the population size used in the simulation. Probabillity distributions are shown with box-whisker plots (black line: mean value, box: inter-quartile range, fences: max and min values). Here, the population size is expressed in terms of the critical population size N^* (Eq.31 in SI Appendix); evolutionary experiments performed with an effective population size $N \simeq N^*$ are governed by effects C.I. and have a high probability ($= \frac{1}{2}$ for $N = N^*$) to display two distinct phases. Taken together, these results show that, when CI is very likely to take place (i.e., for $N \simeq N^*$, cfr SI Appendix) since the anueploids can reach a substantial maximum frequency ($x_a^{Max} > 0.5$), this dynamics could be wrongly interpreted as a fixation of aneuploids followed by loss of the extra chromosome. Model parameters used in the simulations: (i) selection coefficients $\sigma_b = 0.17\,\mathrm{gen}^{-1}$, $\sigma_c = 0.05\,\mathrm{gen}^{-1}$),(ii) mutation rates $\mu_a = 4 * 10^{-4}\mathrm{gen}^{-1}$ and$\mu_m = 2 * 10^{-5}\mathrm{gen}^{-1}$, (iii) population size $N = 2000$ (Panel A), 1500,2000,2500 (Panel B). With this set of model parameters, the critical population size (Eq.31 in SI Appendix) takes the value $N^* = 1983 \simeq 2000$.

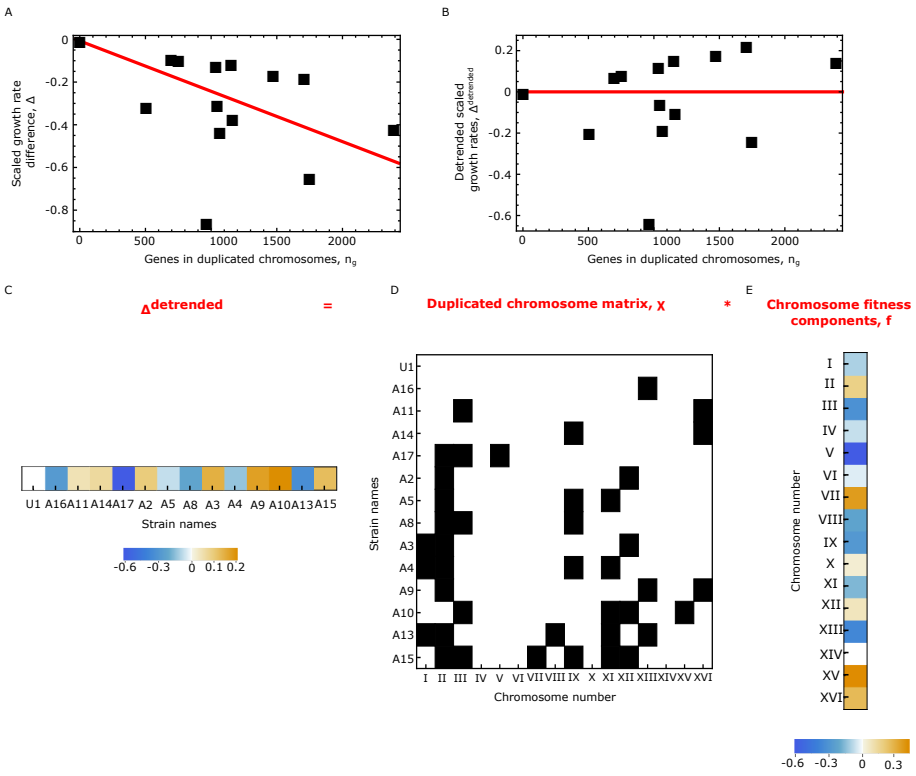

Figure S10: **Detailed procedure of the inference of chromosomal fitness components from scaled growth rate.** We show here a detailed example of the algorithmic and mathematical steps involved in the inference of the fitness components from scaled growth rates from ref. [8]. In this example, we show growth data evaluated at 20°C, for strains with for pl=1 background. Panel A shows the values of the scaled growth rate differences Δ_s (computed according to Eq. 19, Material and Methods) *vs* the number of duplicated genes contained in the corresponding strain (n_g^s). The red line shows the linear decreasing trend for this data, which, in our model, is related to the average fitness cost of a gene contained in a duplicated chromosome. The first step of our inference removes this (genome-wide) linear trend, by detrending the data (according to Eq. 23, Material and Methods). The result of this detrending procedure is shown in panel B, the red line now showing the absence of a residual trend. The set of detrendend data of all the considered strains can be mathematically represented as a vector $\Delta^{\vec{\text{detrendend}}}$, as shown in panel C, where numerical values are now color coded (legend at the bottom of the panel). To infer the chromosome-specific fitness components we decompose this set of values ($\vec{\Delta}^{\text{detrendend}}$) into the product of the duplicated chromosomes χ and the vector of the chromosomal fitness components ($\vec{f}$). The matrix χ, shown in panel D, specifies the set of duplicated chromosomes in a given strain; here duplicated chromosomes are shown in black (corresponding to a numerical value =1) while non-duplicated chromosomes are shown in white (corresponding to a numerical value =0). The vector of the fitness components specifies the contribution of each chromosome to the observed values of the detrended growth rates ($\Delta^{\vec{\text{detrendend}}}$) and is computed according to Eq.23, Material and Methods. Note that the algebraic relation ($\vec{\Delta}^{\text{detrendend}} = \chi * \vec{f}$) implies that the detrended growth rate of a given strain is the sum of the fitness components of its duplicated chromosomes only.

Table S1: **Inferred Values of the growth rates of the aneuploid and euploid strain from ref. [6]**

High Temperature

	Estimate (h^{-1})	Standard Error (h^{-1})
$f_{\text{eu}}^{30°C}$	0.439	0.004
$f_{\text{an}}^{30°C}$	0.415	0.006
$f_{\text{eu}}^{39°C}$	0.253	0.005
$f_{\text{an}}^{39°C}$	0.284	0.003

High Ph

	Estimate (h^{-1})	Standard Error (h^{-1})
$f_{\text{eu}}^{\text{normal ph}}$	0.392	0.005
$f_{\text{an}}^{\text{normal ph}}$	0.345	0.005
$f_{\text{eu}}^{\text{high ph}}$	0.109	0.002
$f_{\text{an}}^{\text{high ph}}$	0.154	0.005

Table S2: **Posterior mean values of the model parameters for the logistic fit $y(t) = \frac{be^{ft}}{1+a(e^{ft}-1)}$ of the growth data from ref. [7]**

Strain	# exceeding genes	f (h^{-1})	a	b	inverse variance $1/\sigma_Y^2$
Eu	0	0.5	0.022	0.22	38
Dis. I	117	0.46	0.026	0.25	48
Dis. II	456	0.46	0.025	0.24	48
Dis. IV	836	0.2	0.053	0.22	169
Dis. VI	139	0.29	0.019	0.22	35
Dis. VIII	321	0.45	0.031	0.28	43
Dis. IX	241	0.47	0.015	0.15	35
Dis. X	398	0.44	0.028	0.24	40
Dis. XI	348	0.42	0.033	0.24	51
Dis. XII	578	0.34	0.02	0.17	29
Dis. XIII	505	0.37	0.026	0.24	48
Dis. XIV	435	0.38	0.031	0.23	61
Dis. XV	597	0.33	0.0095	0.09	55
Dis. XVI	511	0.35	0.012	0.13	25.
Dis. XVI+XI	859	0.27	0.016	0.1	19
Dis. XIV+VIII	756	0.3	0.016	0.1	17
Dis. XI+XV	945	0.21	0.016	0.12	33

Table S3: **Best-fit model parameters for the aneuploid strains abundances data** [10]. Numerical values of the model parameters for the fit of Eq.9 (Main text), to aneuploidy strains abundances data. In (i) we considered the full data-set presented in ref. ([10], where authors aggregated data for aneuploidy strains from 8 independent studies. In (ii) we split the data-set into two subsets, the subset of "natural strains" from ref. [11] and its complement set, in order to evaluate differences between model parameters for the two subsets. Similarly, in (iii) we compared the set of "wild strains" from ref. [10] and its complement set. In (iv) we focused on the subset of natural strains only, which we further partitioned into aneuploid strains with a ploidy>2 background and strains with ploidy=1 and ploidy=2 background.

	Dataset	Normalization, $1/Z$	Effective fitness cost per gene, κ	N_{strains}
(i)	All strains (ref.s [11–18])	24000.	5×10^{-4}	936
(ii)	Natural strains (ref.s [11])	6300.	2.1×10^{-4}	295
	Complement set (ref.s [12–18])	17000.	5.7×10^{-4}	641
(iii)	Wild Strains(ref.s [11, 13])	4400.	2.2×10^{-5}	237
	Complement (ref.s [11, 12, 14–18])	20000.	7.4×10^{-4}	699
(iv)	Natural strains, ploidy >2 (ref [11])	1100.	2.5×10^{-7}	63
	Natural strains, ploidy 1 and 2 (ref [11])	5100.	3.7×10^{-4}	232