Evolutionary trade-off and mutational bias could favor transcriptional over translational divergence within paralog pairs

Simon Aubé[1,2,3,4*], Lou Nielly-Thibault[2,3,4,5], Christian R. Landry[1,2,3,4,5*]

1 Département de Biochimie, de Microbiologie et de Bio-informatique, Faculté des Sciences et de Génie, Université Laval, G1V 0A6, Canada
2 Institut de Biologie Intégrative et des Systèmes, Université Laval, G1V 0A6, Canada
3 PROTEO, Le regroupement québécois de recherche sur la fonction, l'ingénierie et les applications des protéines, Université Laval, G1V 0A6, Canada
4 Centre de Recherche en Données Massives, Université Laval, G1V 0A6, Canada
5 Département de Biologie, Faculté des Sciences et de Génie, Université Laval, G1V 0A6, Canada

* simon.aube.2@ulaval.ca; Christian.Landry@bio.ulaval.ca

Abstract

How changes in the different steps of protein synthesis – transcription, translation and degradation – contributed to differences of protein abundance among genes is not fully understood. Here, we show that yeast paralogous genes are more divergent in transcription than in translation. We explore two causal mechanisms for this predominance of transcriptional divergence: an evolutionary trade-off between the cost and precision of gene expression and a larger mutational target size for transcription. Performing simulations within a minimal model of post-duplication evolution, we find that both mechanisms are consistent with the observed divergence patterns. We also investigate how additional properties of the effects of mutations on gene expression, such as their asymmetry and correlation across levels of regulation, can shape the evolution of duplicates. Our results highlight how general trade-offs in cellular processes and mutation bias can have far-reaching impacts. If transcriptional divergence is favored between paralogs independently of gene-specific functional constraints, it may have important implications for the evolutionary potential of gene duplication and the contribution of this process to the general evolution of gene expression levels.

Introduction

Gene expression level – the steady-state cellular abundance of the corresponding protein – is a fundamental property of genes, as shown by extensive reports of fitness-expression dependencies across organisms and biological functions [1–4]. As such, understanding its evolution is of great biological importance. Gene expression is a multi-step process, involving the synthesis of mRNAs and proteins as well as their active degradation and dilution during cell division. Accordingly, variation in protein abundance among genes within or between species can arise from changes at multiple levels. While many studies have described the crucial role of mutations within genes, their regulatory sequences and their regulators in generating expression variation, most reports focused on one of these aspects, for instance on the transcriptional component [5]. How each of these levels of regulation changes during evolution, independently or jointly, remains to be addressed.

One powerful context in which the evolution of gene expression can be examined is that of gene duplication. Notably, gene duplication is frequently associated with the divergence of the resulting paralogs into two functionally distinct genes through processes known as neofunctionalization and subfunctionalization, respectively involving the acquisition of new function(s) [6] and the partitioning

of the ancestral function [7]. This divergence often includes protein abundance changes. Post-duplication expression reduction has for instance been reported in paralog pairs [8] and shown to occur in transcription and in translation [9]. Moreover, a compensatory drift of expression levels may occur, allowing both gene copies to diverge while maintaining a constant cumulative protein abundance [10, 11]. Gene duplication, through its high frequency – potentially above that of single-nucleotide mutations [12, 13] – and far-reaching consequences, may thus shape the evolution of gene expression levels genome-wide.

Some features of paralogous genes additionally simplify the study of the evolution of expression in transcription and translation. Because most duplicate pairs were initially two identical copies of the same gene, both paralogs were likely expressed at a similar level. Their relative expression divergence can thus be estimated directly from the relative distance between them, without prior reconstruction of the ancestral state and precise evolutionary trajectories. If done for many gene pairs at the levels of transcription and translation, such inference of relative expression divergence can be especially informative. Since a given change in protein abundance can be obtained from an infinity of combinations of transcriptional and translational divergences, any consistent evolutionary pattern across numbers of paralog pairs would be telling about the underlying mutational process as well as the selective pressures potentially involved. In addition, since gene duplicates share the same cellular environment, their divergence of expression by definition involves an important component of cis-regulatory changes, which we can infer to be changes in the properties of the genes, mRNA and proteins themselves and their regulatory elements.

In order to elucidate how duplicated genes diverge in transcription and translation – and thereby more fully understand the evolution of gene expression levels –, we leveraged a recently published set of transcription and translation rates for 4440 genes of the budding yeast *Saccharomyces cerevisiae* [14]. This organism is a well-recognized model for the study of gene duplication, notably because it is descended from a WGD event which occurred ~ 100 million years ago [15]. We first assessed differences in transcription and translation rates within paralog pairs, which revealed that expression mostly diverged at the transcriptional level. This led us to explore why divergence would accumulate more at this level of regulation. We hypothesized that two mechanisms could explain this predominance of transcription changes: an evolutionary trade-off in the optimization of expression levels and a larger mutational target size for transcription than for translation. Using *in silico* evolution, we show that both hypotheses are consistent with the observed divergence patterns. We conclude by stressing the importance of measuring the mutational target sizes and other mutational parameters of real genes in order to be able to further support one model or the other.

Results

Yeast duplicates mostly diverged in transcription

We first examined the transcriptional and translational divergence within yeast paralog pairs using transcription rates β_m and translation rates β_p – respectively in mRNAs per hour and in proteins per mRNA per hour – from Hausser et al. (2019) [14]. Under this formalism, protein abundance is proportional to the product of the two rates, such that they both contribute equally to overall expression and relative changes at the two levels are directly comparable.

$$fold_{dupli} = log_2 \left(\frac{max\left(\theta_1, \theta_2\right)}{min\left(\theta_1, \theta_2\right)} \right) \tag{1}$$

Among the 4440 genes for which these rates were estimated, we identified 409 high-confidence paralog pairs, each derived from a whole-genome duplication (WGD; $n = 245$) or a single small-scale duplication event (SSD; $n = 164$). We assessed the contributions of transcriptional and translational changes by computing the magnitude of relative divergence in transcription and translation as in Eq (1), where θ represents any property (e.g. transcription or translation rates) of paralogs 1 and 2 within a pair.

The distributions of these two relative measures for all paralog pairs reveal that expression mostly diverged transcriptionally (Fig 1A). The median magnitude of relative divergence is 1.5 times larger in transcription than in translation, with only slight differences depending on duplication type (1.45x for WGD and 1.48x for SSD). In addition, about 70% (74.3% for WGD and 62.2% for SSD) of paralog pairs are more divergent in transcription rates than in translation rates. Using normalized differences instead of log_2 fold-changes (Methods), leads to the same conclusions (S1A Fig). We also correlated the relative magnitudes of transcriptional and translational divergence across gene pairs, which revealed a weak but significant positive association (Fig 1B). Paralog couples which diverged more in transcription thus also tend to have accumulated more changes in translation. This may reflect the action of gene-specific selective constraints on protein abundance or the existence of a correlation between transcriptional and translational mutations (see below). Repeating the same analysis without first defining log_2 fold-changes of expression as strictly positive as in Eq 1 again shows that transcriptional and translational changes mostly occurred in the same direction within a paralog pair (Fig 1C).

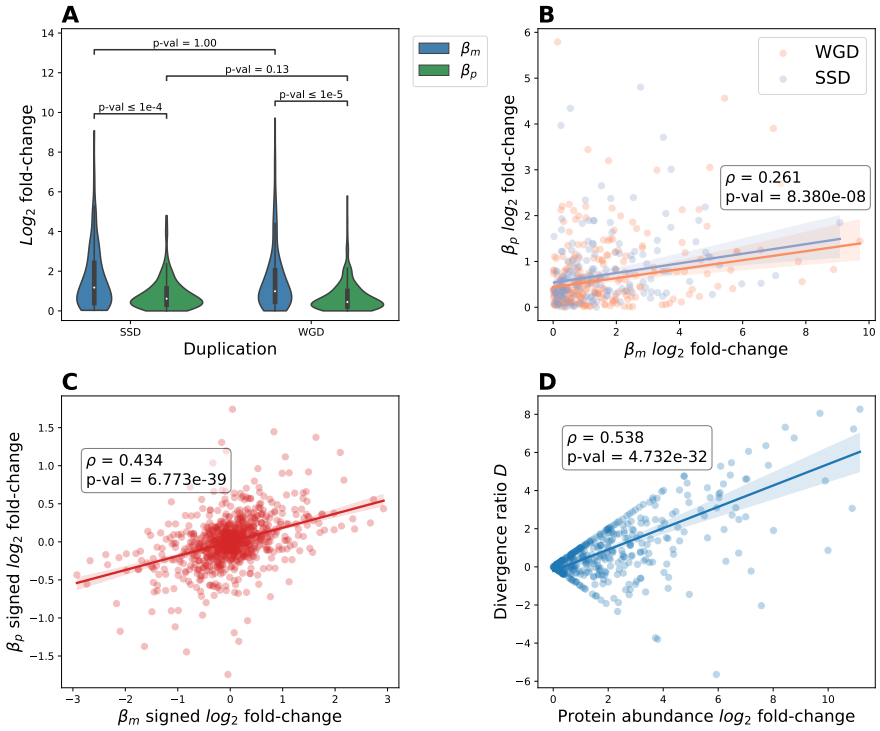

Fig 1. Gene expression divergence of duplicated genes is greater in transcription than in translation. (A) Distributions of relative divergence for *S. cerevisiae* paralog pairs according to the associated mechanism of duplication (245 WGD pairs; 164 SSD pairs). P-values from Mann-Whitney-Wilcoxon two-sided tests are shown. (B) Correlation between the magnitudes of relative transcriptional and translational divergence across pairs. Spearman's ρ and p-value are shown. (C) Correlations between signed fold-changes of transcription and translation rates for all paralogs. Signed log_2 fold-changes were calculated in the two possible orientations for each gene pair and the correlation was computed on the resulting duplicated dataset ($n = 818$). (D) Correlation between the divergence ratio D and the relative divergence in protein abundance for all paralogs ($n = 409$). The triangular structure of the resulting scatter plot is an artifact of our previous definition of transcriptional, translational and protein abundance relative divergences as strictly positive in Eq 1. When these ratios are computed in one constant orientation for the whole dataset – such that their log-transformed values can be negative – and D is calculated accordingly, this structure disappears while the correlation remains of a similar magnitude (S1C Fig).

These evolutionary patterns are strikingly similar to what was recently shown in model plants

Arabidopsis thaliana and *Zea mays* [16], where both WGD- and SSD-derived paralogs diverged more in mRNA abundance than in translation efficiency (analogous to transcription rate and translation rate, respectively). One notable difference however is that this study reported a predominance of compensatory changes at the transcriptional and translational levels, which disagrees with the positive correlation that we observed between signed divergences in transcription and translation (Fig 1C). This discrepancy may be related to differences in selection efficiency, as the effective population size of *A. thaliana* is one to two orders of magnitude lower than that of yeasts [17–19]. Plants also are multicellular, which means that paralogs could diverge along other dimensions that we are not capturing in yeast. Finally, gene regulation may have additional layers of complexity in these species.

In order to better describe the regulatory evolution of yeast duplicates, we also introduce a divergence ratio D, which quantifies the transcriptional or translational bias in the expression divergence of a gene pair. Positive values indicate a preponderance of transcription changes, while negative ones are instead associated with a mostly translational expression divergence.

$$D = log_2 \left(\frac{max\left(\beta_{m1}, \beta_{m2}\right) / min\left(\beta_{m1}, \beta_{m2}\right)}{max\left(\beta_{p1}, \beta_{p2}\right) / min\left(\beta_{p1}, \beta_{p2}\right)} \right) \tag{2}$$

We computed D and its correlation with the magnitude of protein abundance divergence. This revealed a significant positive association (Fig 1D), meaning that the more divergent paralog pairs also tend to show a greater bias towards transcriptional changes. This relationship is likely due to the greater variance of β_m log_2 fold-changes, and thus illustrates how transcriptional changes drive expression divergence. As the transcription and translation rates of both paralogs are used in the calculation of their pair's relative protein abundance divergence and divergence ratio D, we also assessed whether the observed correlation between these two properties could be spurious. A permutation analysis involving the sampling of randomized paralog pairs from our set of 818 individual *S. cerevisiae* paralogs revealed the correlation shown in Fig 1D to be significant with p-val< 0.0001 (S1D Fig).

A minimal model of post-duplication expression evolution

We sought to explain why yeast paralogs mostly diverged transcriptionally and considered two hypotheses (Fig 2A). First, transcriptional mutations could accumulate faster if they had less deleterious fitness effects than translational ones. Such a discrepancy may emerge through a recently described evolutionary trade-off [14], according to which corresponding changes in transcription and translation are not fully equivalent. Briefly, it postulates that the cost and precision of gene expression need to be optimized for each gene and are set by the relative contributions of transcription and translation. A larger transcriptional contribution reduces the magnitude of stochastic fluctuations in protein abundance [20, 21] – or expression noise – and may be beneficial, but simultaneously incurs additional metabolic costs through the synthesis of more mRNA molecules [14, 22, 23]. Conversely, an increased translational contribution results in a noisier expression at a lower cost, since fewer transcripts are synthesized. A higher reliance on translation indeed has no impact on the cost of protein synthesis, because the latter depends on how much protein is produced, regardless of whether it is from the translation of many mRNAs at low rates or few mRNAs at high rates. Changing protein abundance by changing β_m or β_p is therefore not equivalent within this framework.

Alternatively, our observation could be explained in a more parsimonious manner – without any direct involvement of selection in the preferential accumulation of transcriptional changes – if transcription was more likely to be altered by mutations. Our second hypothesis is thus that transcription has a larger mutational target size than translation (Fig 2A), meaning that mutations acting on this trait would be more frequent or have larger effects or both. Under neutral evolution or even under selection to maintain protein abundance, more changes would thereby accumulate at the transcriptional level.

To test these hypotheses, we defined a minimal model of post-duplication evolution based on previous work [8, 10]. This model considers that two paralogs are identical and express the same

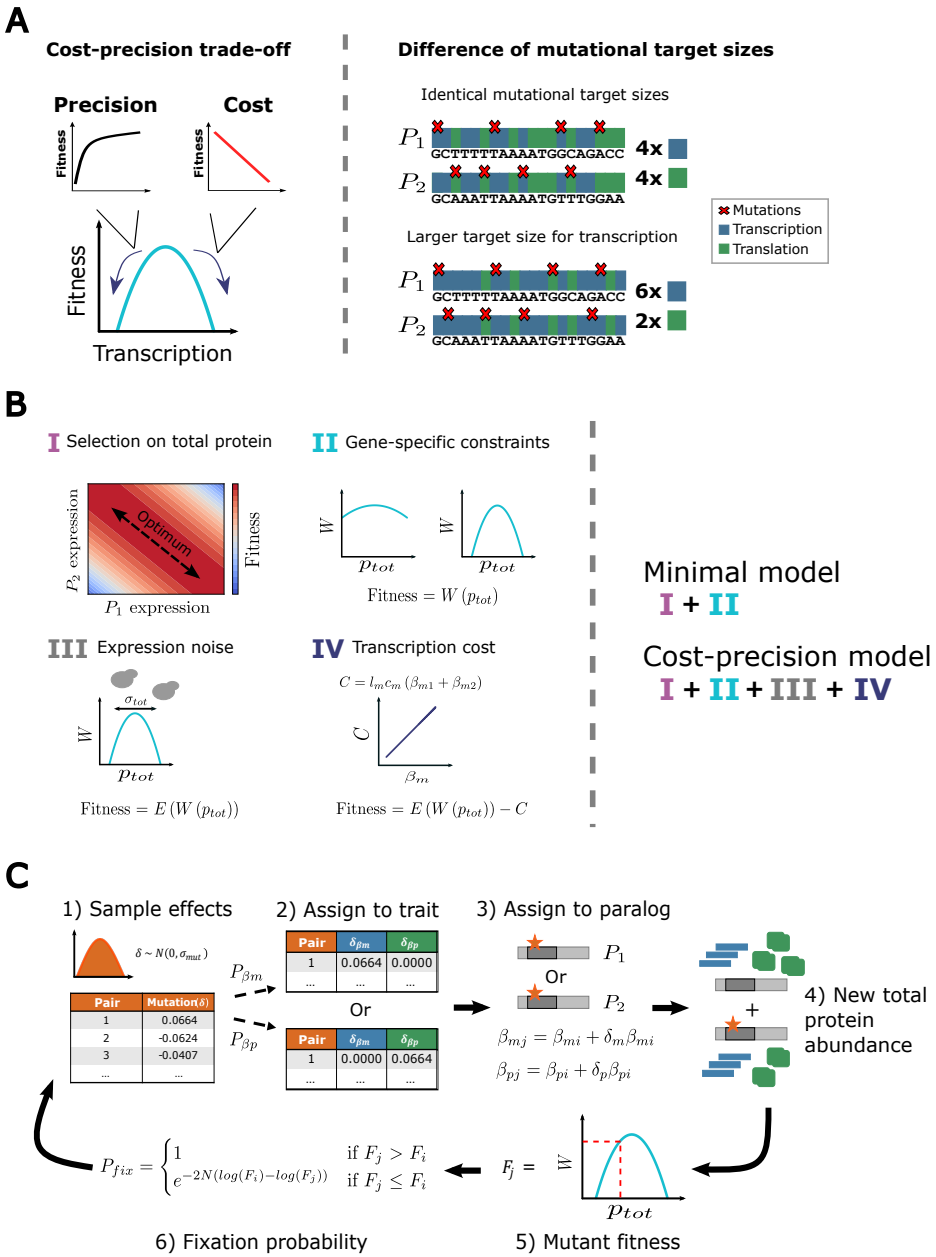

Fig 2. Overview of the modeling and simulation approach. (A) Two hypotheses are considered: an evolutionary trade-off between the cost and precision of gene expression (left) or a larger mutational target size for transcription than for translation (right). (B) A minimal model of post-duplication evolution is defined under the assumption that selection acts to maintain the cumulative protein abundance of two paralogs. It is complexified by the addition of cost-precision constraints [14] to test the hypothesized evolutionary trade-off. (C) The expression divergence of paralogs is simulated by sequential fixation, through multiple rounds of mutation-selection. The process is repeated as long as a realistic level of protein abundance divergence has not been reached (Methods).

protein at the timing of duplication and that selection acts on the total expression of the protein, which has an optimal value. Accordingly, we model the evolution of the expression levels of two paralogous genes across a landscape of diagonal fitness isoclines where the optimum is along a central diagonal of constant cumulative expression (Fig 2B I). Such a landscape is obtained from a parabolic function of fitness according to the cumulative protein abundance p_{tot} of both paralogs. For our minimal model to be applicable to the complete yeast genome, a family of functions with varying curvatures – taken from a distribution inferred from Hausser et al. (2019) [14] (see Methods) – is defined so as to include gene-specific selective constraints (Fig 2B II). By only allowing the evolution of protein abundance, our model also assumes that the expression divergence of two paralogs is independent from their functional divergence. This is of course unlikely to be completely true, but this simplifying assumption appears valid within the scope of this work. Our analyses indeed show at worst a weak relationship between the divergence ratio D and various proxies of functional divergence, meaning that the process generating a dominance of transcriptional divergence is largely independent from the accumulation of changes of function within a pair (S2 Fig).

To implement the cost-precision trade-off within this minimal model, the latter was complexified in two ways. First, expression noise was taken into account by computing the mean fitness of a population expressing two paralogs at a mean cumulative level P_{tot} with standard deviation σ_{tot}, which are themselves calculated from the protein abundances and transcription rates [14] (Fig 2B III; Methods). Second, a penalty C increasing linearly with the total number of transcribed nucleotides was applied to fitness to account for the cost of transcription [14] (Fig 2B IV; Methods).

We performed *in silico* evolution of simulated paralog pairs under both the minimal model and its cost-precision variant. These simulations were carried out according to a sequential fixation approach [24]. Each successive mutation was instantaneously brought to fixation or rejected.

To initialize each simulation run, random singleton genes were first generated by sampling protein abundances and fitness function curvatures from the complete yeast dataset [14] and optimizing their transcription rates β_m and translation rates β_p according to the cost-precision trade-off (Methods). Then, each singleton was duplicated into two paralogs P_1 and P_2 which both have the ancestral β_m and β_p rates. We assumed that a duplication event causes a doubling of total transcriptional output without affecting the translation of individual transcripts, which is realistic since the translation rates of most mRNA do not change upon gene copy number increase [25]. Following the duplication of any random singleton, an optimum of cumulative protein abundance p_{tot} is set according to the product of the ancestral protein abundance by a constant Δ_{opt}, such that the expression doubling is not necessarily optimal but the immediate loss of a newly created paralog is deleterious. The lowest such constant which prevents negative post-duplication fitness in genes with the lowest possible tolerance to expression fluctuations was used (1.86; Methods).

Once the paralog pairs had been generated, they were all subjected to successive mutation-selection rounds so that their transcription and translation rates could evolve, while maintaining an approximately constant cumulative protein abundance (S1 File). These were performed as follows (Fig 2C). Briefly, one relative mutational effect is first sampled for each pair from a normal distribution of mean 0 and standard deviation σ_{mut}. This relative expression change is then randomly assigned to transcription or translation according to probabilities $P_{\beta m}$ and $P_{\beta p}$, which represent the relative mutational target sizes of the two traits. The transcriptional or translational mutation is next applied randomly to paralog P_1 or P_2, increasing or decreasing its expression by a percentage. Mutations affecting both duplicates simultaneously – for instance changes in *trans* – are therefore ignored, because they would not directly result in any expression divergence, and all randomly-selected mutations can be considered as *cis*-acting within this framework. Once the mutational effect has been applied, a new cumulative protein abundance can be computed, from which a mutant fitness F_j is then calculated either according to the minimal model or its cost-precision version. From the resulting F_j and the ancestral fitness F_i, a fixation probability is computed using a modified Metropolis criterion [26], according to which the attempted mutation reaches fixation or is lost. We note that this modified criterion can be adjusted for different levels of selection efficacy with a parameter N analogous to the effective size of the modeled population. Any mutation taking β_m or β_p to 0 or lower

is in addition automatically rejected at this step.

This complete mutation-selection process was repeated iteratively during each simulation run, until the median relative protein abundance divergence (Eq 1) for the simulated set of duplicate pairs was not significantly different from the empirical value observed in yeast paralogs (Methods).

The cost-precision trade-off is sufficient to promote transcriptional divergence

We performed a small-scale 'mock' simulation of 50 randomly generated duplicate pairs according to the cost-precision implementation of our minimal model. Transcription and translation were assumed to have equal mutational target sizes ($P_{\beta m} = P_{\beta p}$), the standard deviation of mutational effects was set to a small magnitude ($\sigma_{mut} = 0.025$) and a scenario of high selection efficacy was considered ($N = 1 \times 10^6$).

This simulation revealed a clear bias towards transcriptional changes, with the relative divergence in transcription accounting for almost all the total variation of protein abundance (Fig 3A). After almost 5000 mutation-selection rounds, the median relative divergence was nearly twice larger for β_m than for β_p across the 50 simulated paralog pairs. The resulting evolutionary trajectories moreover highlight that expression divergence was driven by transcriptional changes, as the most transcribed paralog is almost always associated with the highest protein abundance, while the same is not true for the most translated one (Fig 3C).

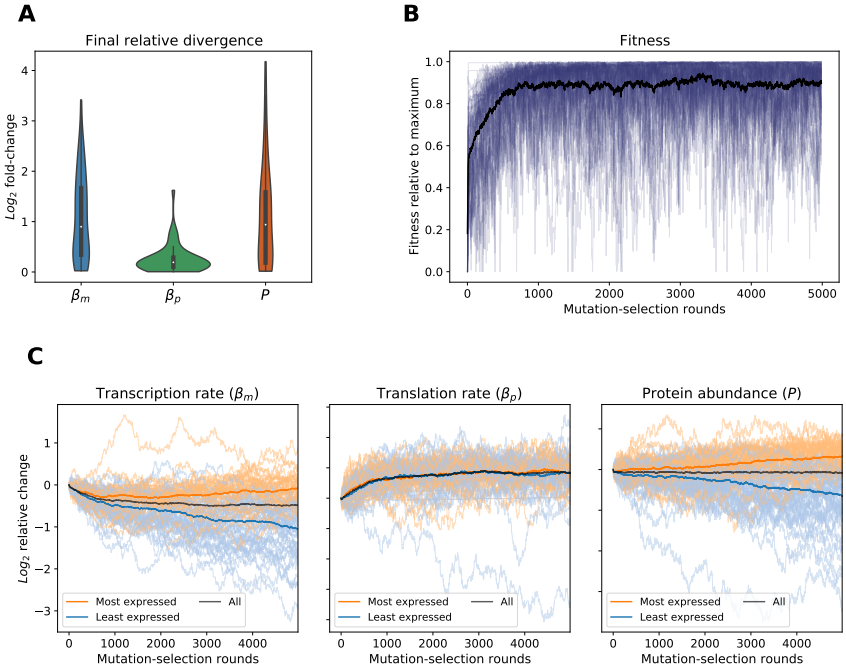

Fig 3. The cost-precision trade-off favors transcriptional divergence. (A) Distributions of the final logfold relative divergence in transcription, translation and protein abundance for the simulation of 50 randomly generated paralog pairs. The simulation was stopped when it matched the protein abundance divergence of WGD-derived duplicates (Methods). (B) Fitness associated with each of the 50 duplicate pairs throughout the simulation. The mean value at any time point is shown by the black line. Fitness is scaled individually for each gene pair between its minimum (0) and maximum (1) values reached during the simulation. (C) Full evolutionary trajectories of the 50 simulated paralog pairs in transcription rate, translation rate and protein abundance. All three features are shown as log-scaled relative changes from the ancestral value for each duplicate pair and the darker lines represent the means through time.

Interestingly, these patterns do not arise because the divergence of transcription levels is itself beneficial under the cost-precision trade-off, as illustrated by the fitness plateau observed from round 1000, while transcription rates are actively diverging (Fig 3B). Indeed, because expression noise scales according with both protein abundance and transcription rate (Methods), variance in the cumulative expression of a duplicated gene depends on the total β_m and expression level of its two copies. Similarly, the cost of transcription only depends on the total number of mRNAs synthesized, such that the distribution of transcriptional contributions between two paralogs has no impact on fitness. Rather, it is the ratio of cumulative translation and transcription rates which dictates fitness under cost-precision constraints, as clearly shown by the early adaptation phase of the simulated evolutionary trajectories (Fig 3B). After an initial reduction of cumulative protein abundance to reach the new post-duplication expression optimum, the transcription of both duplicates is decreased while their translation is increased (Fig 3C), to rebalance the $\frac{\beta_p}{\beta_m}$ ratio [14] of the pair. Following this first phase of post-duplication adaptation, compensatory drift of cumulative protein abundance takes place [10, 11], but interactions between transcriptional and translational mutations under the cost-precision trade-off bias it almost exclusively towards transcription divergence.

This first simulation thus shows that the trade-off between the cost and the precision of gene expression can induce a mostly transcriptional expression divergence between two paralogs, potentially creating evolutionary patterns as observed in yeast.

Selecting a biologically plausible parameter space

The cost-precision trade-off is sufficient to favor the transcriptional divergence of duplicated genes, but it might still not be the best explanation for the evolutionary patterns of yeast paralogs. To rigorously compare our two competing hypotheses, we performed series of large-scale simulations under biologically realistic assumptions. All the corresponding *in silico* evolution experiments were carried out in three replicates of 2000 randomly generated paralog pairs. They also focused solely on WGD-derived duplicates as reference set, since the minimal model of post-duplication evolution that we consider has been proposed previously in the context of whole-genome duplications [10] and its applicability to other duplication events is unclear. Ensuring the biological plausibility of these simulations additionally required a careful examination of two important parameters within our model – the efficacy of selection N and the standard deviation of mutational effects σ_{mut} –, which are not known precisely.

Many estimates of the effective population size of *Saccharomyces* yeasts, which may correspond to the N parameter of our evolutionary algorithm [26], are available [18, 19], but the relevance of such historical values to the immediate post-WGD context is unclear. Identifying a most realistic N is then hazardous at best, and we instead chose to consider two scenarios, respectively of high ($N = 1 \times 10^6$) and reduced ($N = 1 \times 10^5$) efficacy of selection, and compare our two hypotheses within both contexts.

Similarly, although the distribution of mutational effects on expression level has been studied in budding yeast, only limited information is available on the effect of mutations – and especially *cis*-occurring ones – on protein abundance. Previous studies focused on only a small subset of genes [27, 28], did not differentiate between *cis* and *trans* mutations [28, 29], assessed too few mutations [29] or were limited to substitutions occurring within a short segment of the promoter sequence [30]. Thus, instead of arbitrarily choosing a σ_{mut}, we performed simulations across a range of standard deviations, from 0.01 to 0.35, to identify the most biologically plausible value.

Because our second hypothesis only stipulates that the regulation of transcription may have a larger mutational target size than that of translation without specifying the magnitude of this difference, testing it requires using various relative probabilities of transcriptional and translational mutations, P_{β_m} and P_{β_p} (Fig 2C). To make it more robust, the identification of the best-fitting σ_{mut} was therefore combined with this screening into a grid search, which was performed separately for the two scenarios of selection efficacy. In each instance, the most plausible standard deviation of mutational effects was defined as the one resulting in the best overall fit to the empirical distributions of relative divergence – at the levels of transcription, translation, protein abundance and divergence ratio D – for any combination of model (minimal or cost-precision) and relative mutational target sizes (Methods).

This approach identified best σ_{mut} values of 0.025 and 0.100, respectively under high ($N = 1 \times 10^6$) and reduced ($N = 1 \times 10^5$) selection efficacies (S3 Fig). Interestingly, the value obtained for $N = 1 \times 10^6$ is highly consistent with a previous experimental characterization of *cis*-regulatory mutations in the yeast *TDH3* promoter [27], which may however not be fully representative of the typical *S. cerevisiae* gene [28].

A difference of mutational target sizes may better explain the observed divergence patterns

We assessed the extent to which our minimal model and its cost-precision variant could replicate the main features of the divergence patterns of yeast paralogs (see Fig 1): 1) a predominance of transcriptional changes, 2) a weak positive correlation between the magnitudes of relative divergence in transcription and in translation, 3) a high frequency of amplifying changes at the two regulatory levels and 4) a tendency for the most diverged gene pairs to show the strongest bias towards transcriptional changes. We note that this assessment is not limited to the two alternative hypotheses in isolation. Any cost-precision simulation performed with unequal probabilities of transcriptional and translational mutations indeed tests a combination of the two mechanisms.

We first compared empirical and simulated distributions of relative divergence in transcription and translation rates, protein abundance and divergence ratio D across all tested combinations of parameters. The Kolmogorov-Smirnov (KS) statistic was used to quantify the distance between distributions and a mean value was computed for each replicate simulation. The resulting means of KS statistics reveal the contrasting performance of both models depending on parameters choice (Fig 4A). When a high efficacy of selection is assumed – intuitively more likely for *Saccharomyces yeasts* –, only the minimal model can reasonably replicate the empirical patterns of divergence. This is exclusively the case when a higher probability of mutations affecting β_m is assumed, supporting the hypothesis of a larger mutational target size for transcription. The poor performance of the cost-precision model is almost entirely due to its inability to produce a realistic translational divergence (S4A-B Fig). We thus performed additional simulations with $P_{\beta p} > P_{\beta m}$, which however did not result in a better fit to the empirical expression changes (S5 Fig). In contrast, following a reduction of the efficacy of selection, both models perform much more similarly (Fig 4A). This highlights how both hypothesized mechanisms might have shaped the expression divergence of yeast paralogs, although the mutational target size difference hypothesis is more robust to the postulated efficiency of selection. Interestingly, the best agreement with the real patterns of evolution is obtained for a relatively modest difference of mutation probabilities.

The comparison between the magnitudes of relative divergence in transcription and in translation within simulated gene pairs reveals similarities in performance between our two models, with little influence of selection efficacy (Fig 4B). In both cases, most replicate simulations generate realistic divergence correlations which fall within the 95% confidence interval of the empirical correlation coefficient, although it does not always coincide with the replication of the true distributions of relative divergence (Fig 4A). Because both models are able to reproduce the observed relationship between the magnitudes of transcriptional and translational changes, this further highlights the plausibility of the two alternative hypotheses. As previously, these results however provide slightly more support for the hypothesis of a larger mutational target size for transcription, since all simulated correlations obtained from the minimal model are realistic across the two levels of selection efficacy – and their distribution even appears centered on the empirical value when $N = 1 \times 10^6$.

If the two models accurately described the evolution of duplicated genes, they would also replicate the signed divergence correlations between relative changes in transcription and in translation (Fig 1C). We computed this correlation across the diverged paralog pairs obtained from all simulations and compared the resulting values to the relationship observed empirically. As shown in Fig 4C, our models are unable to replicate this feature, and no simulation generated a realistic correlation within the 95% confidence interval of the empirical value. Interestingly, even the minimal model of post-duplication evolution is associated with strictly positive correlations, contrary to the intuitive expectation of a

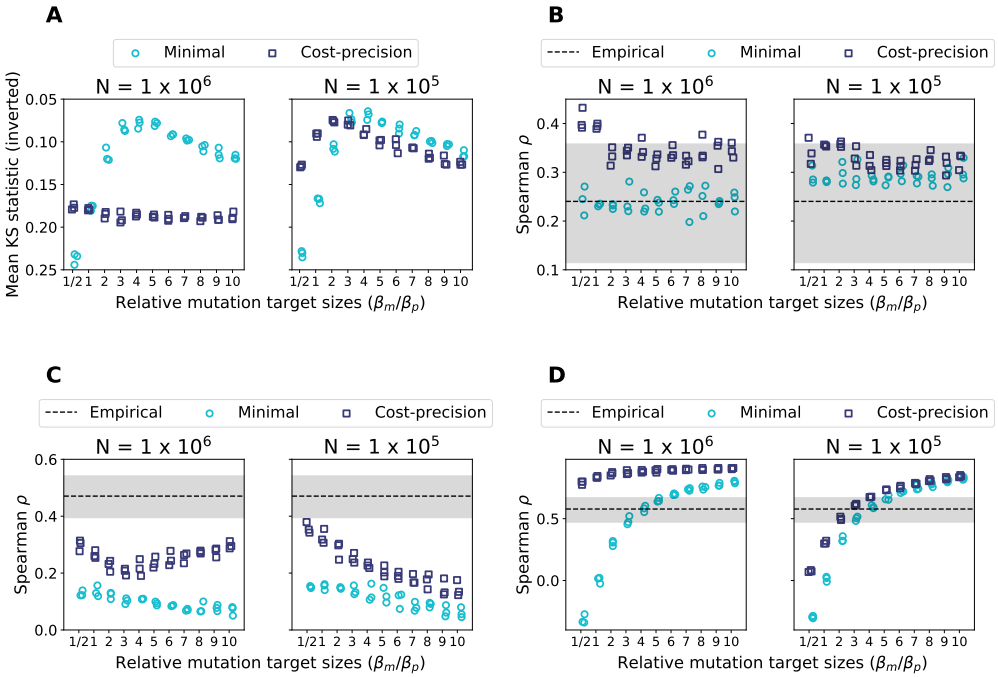

Fig 4. A larger mutational target size for transcription more robustly replicates the different features of the expression divergence of paralogs across levels of selection efficacy. (A) Replication of the distributions of relative divergence in transcription rate, translation rate and protein abundance as well as divergence ratio D following simulations of randomly generated paralog pairs, as shown by the mean of the KS statistics obtained from the pairwise comparisons of all the corresponding empirical and simulated distributions. (B) Correlations between the magnitudes of transcriptional and translational divergences obtained from the *in silico* evolution of duplicate pairs according to the minimal and cost-precision models. The shaded area shows the 95% confidence interval on the empirical correlation observed with *S. cerevisiae* paralogs. (C) Correlations between signed divergences in transcription and translation rates resulting from simulations of expression divergence according to both studied models. The shaded area represents the 95% confidence interval on the empirical correlation in yeast. (D) Correlations between the divergence ratio D and the relative divergence of protein abundance after simulations. The empirical correlation and its 95% confidence interval are shown as the shaded area. Three replicate simulations of 2000 randomly generated paralog pairs were made for each combination of relative mutation probability, selection efficacy N and model. Each simulation was stopped according to the protein abundance divergence of WGD-derived duplicates and evaluated by comparison with the evolutionary patterns observed among these paralogs. A larger relative probability of transcriptional mutations is used to represent a larger mutational target size for transcription.

compensatory drift of expression at both levels under selection to maintain cumulative protein abundance [10, 11]. This may reveal an effect of the postulated post-duplication expression optimum on the divergence correlations. We thus performed additional simulations while varying this optimum, to consider the alternative possibilities that the duplication-induced expression doubling was perfectly optimal or did not sufficiently increase the expression level. These showed a limited dependency of the signed divergence correlation on the posited expression optimum (S6B Fig), which did not invalidate the conclusion that our framework does not realistically replicate this correlation. We note that the effect of the post-duplication optimum was stronger for the correlation between the magnitudes of relative divergence (Fig 4B; S6A Fig), but still insufficient to invalidate our previous conclusion on the reproduction of this feature.

We finally computed the correlation between the divergence ratio D and the relative protein abundance divergence at the end of *in silico* evolution for all simulation runs and compared it to its empirical value (Fig 1D). This analysis reveals a second important disparity between the two alternative evolutionary models. The minimal framework can generate realistic divergence bias correlations for the two levels of selection efficacy, while the cost-precision model only does so when $N = 1 \times 10^5$ (Fig 4D). Importantly, such realistic values coincide with the best reproduction of the empirical divergence patterns – when mutations are 3-5 (minimal model) or 2-3 (cost-precision model) times more likely to affect transcription than translation rates – in both selection scenarios.

These comparative analyses show that both mechanisms are realistic explanations. They however do not identify one definite cause, as each mechanism has limitations. Both are unable to replicate one of the features of the empirical expression divergence, while the cost-precision trade-off is highly sensitive to variations in selection efficacy. Our simulations additionally indicate that a faster neutral rate of evolution for transcription than translation might be needed even if cost-precision constraints were involved. The best performance of the cost-precision model is indeed obtained when it is combined with unequal probabilities of transcriptional and translational mutations. Moreover, the hypothesis of a difference of mutational target sizes arguably emerges as the preferred explanation, since it results in the best fit and is more robust to assumptions about the efficacy of selection. Although we do not know what these relative mutational target sizes are in reality, the fact that the best agreement with our observations is obtained for a modest difference of target sizes means that the bias need not be of a large magnitude to impact evolution.

Realistic mutational constraints improve the replication of expression divergence patterns

A larger mutational target size for transcription can accurately replicate three out of four features of the expression divergence of yeast paralogs. This is however under assumptions that may be too simple, more specifically that mutations independently affect transcription and translation rates, and that all types of expression changes are equally likely.

A given mutation in the transcribed region of a gene may for instance act on transcription and translation by simultaneously affecting mRNA stability and translation efficiency [31, 32]. In addition, mutations at one regulatory level may be associated with amplifying or buffering regulatory changes at the other – as suggested by stress responses [33–35]. Random mutations may also have asymmetrical effects on expression level, as recently shown for yeast genes [28]. Recent work additionally predicted that mutations increasing expression are rarer for highly expressed promoters – and vice-versa for lowly expressed ones [30]. Various constraints on the effects of mutations can thus create correlations between transcriptional and translational changes or bias evolution, and we hypothesized that some of them might allow our minimal model to generate realistic signed divergence correlations (Fig 1C; Fig 4C).

To test this hypothesis, we made two new versions of our simulation model, respectively implementing an asymmetry in the distribution of mutational effects and a correlation between the transcriptional and translational effects of mutations. This second addition to the model notably allows mutations to act on both β_m and β_p at once, in a potentially correlated manner.

Mutation asymmetry was implemented using a skew normal distribution of mutational effects with skewness parameter $\alpha \neq 0$, while correlations between transcriptional and translational mutations were added using a bivariate normal distribution of mutational effects. The latter modification also meant that relative mutational target sizes could not be modeled as mutation probabilities, since each mutation now affected transcription and translation. Instead, they were modeled as differences in mean mutational effects and the standard deviations of transcriptional and translational effects were scaled according to the ratio of target sizes. The precise standard deviations were chosen to obtain the same mean change in protein abundance per mutation as with a given reference σ_{mut} within our standard framework (Methods). Because this represents a significant departure from our previous approach, another grid search was performed in order to identify the best-fitting reference σ_{mut} under each selection efficacy scenarios when using bivariate mutational effects (S7 Fig).

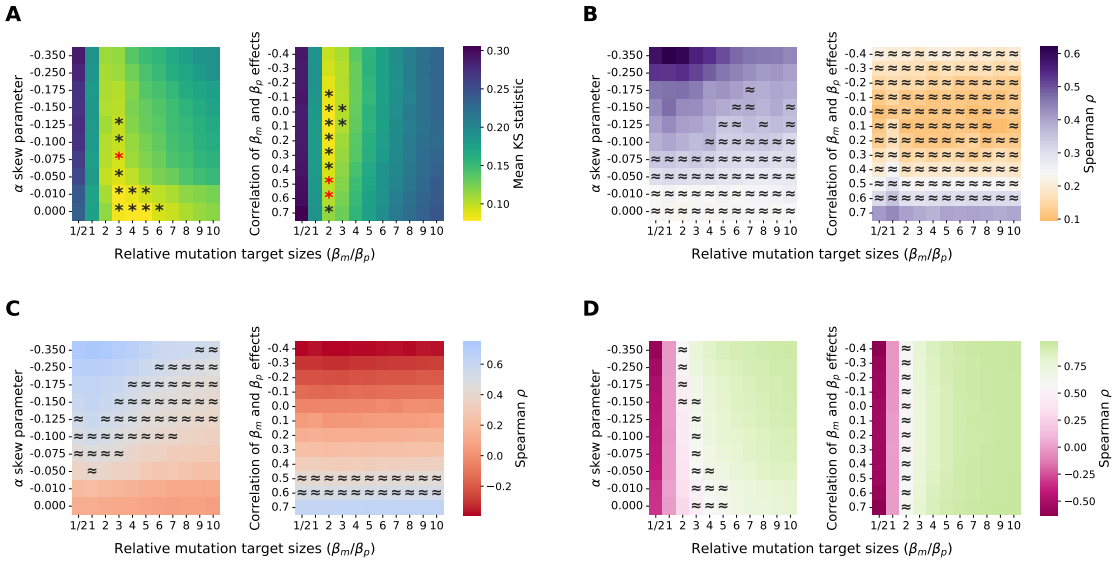

Fig 5. Additional constraints on mutational effects allow the minimal model to simultaneously replicate all four features of the divergence of paralogs. Results shown for three replicate simulations of 2000 randomly generated paralog pairs performed under the minimal model with the assumption of $N = 1 \times 10^6$ and stopped according to the protein abundance divergence of yeast WGD-derived paralogs (A) Average replication of the four distributions of relative divergence (transcription, translation, protein abundance and divergence ratio) for a range of mutational effects distribution asymmetry (left) or correlations between the effects of transcriptional and translational mutations (right). Asterisks identify instances where all four magnitudes of relative divergence are realistic for at least one of three replicate simulations (p-val > 0.05, Mood's median test). A red asterisk indicates that this coincides with the recapitulation of all three correlations studied in the subsequent panels. (B, C, D) Average final correlation across the three replicate simulations between 1) the magnitudes of transcriptional and translational divergences (B), 2) the signed divergences in transcription and translation (C) as well as 3) the divergence ratio D and the magnitude of protein abundance divergence (D) across the same ranges of mutational effects distribution asymmetry or correlations between the effects of transcriptional and translational mutations. In each case, $\approx$ designates parameter combinations where a correlation coefficient within the 95% confidence interval of the empirical value was obtained for at least one replicate simulation.

Simulations were then performed as previously across ranges of distribution asymmetry and correlation of mutational effects (Methods). While both negative and positive correlations between transcriptional and translational effects were assayed, only negative skewness values – biasing mutations towards a decrease of expression – were used, since a positive skew lengthened simulations too much by impeding the initial protein abundance reduction. We particularly focused on the minimal model in a context of high selection efficacy (N $= 1 \times 10^6$), an intuitively likely scenario for yeasts which was however unable to replicate the correlation between the signed relative divergences in transcription and in translation. Both modifications to the mutational process had a strong impact on this correlation and even brought it within the 95% confidence interval of the empirical value (Fig 5C). More importantly, they did not preclude the replication of the other features of the expression divergence of *S. cerevisiae* paralogs, as such rescue of the correlation of signed relative divergence coincides with the obtention of realistic coefficients for the two other studied correlations (Fig 5B, D). Both attempts at complexifying the mutational process however resulted in a reduced agreement between the empirical and simulated divergence distributions, as shown by elevated grand means of KS statistics (Fig 5A). Yet, there are a few instances where the replication of the observed correlations co-occurs with changes in transcription, translation and protein abundance as well as divergence ratio

which are of a realistic magnitude, as evidenced by the absence of significant differences according to Mood's median test (Fig 5A). A larger mutational target size for transcription than for translation is thus fully compatible with the mostly transcriptional divergence of yeast paralogs within our minimal model, at least when potentially realistic mutational biases are added. This is furthermore also the case when the efficacy of selection is reduced (S8 Fig). In contrast, the alternative mechanism of the cost-precision trade-off is only associated once with an accurate replication of all four features, which occurs under the more favorable scenario of $N = 1 \times 10^5$ (S9 Fig).

Discussion

Our analysis of published expression data [14] suggests that transcriptional changes played a greater role than translational ones in the divergence of paralogs in the yeast *S. cerevisiae*. Whether this is a general feature of duplicated genes evolution remains to be fully investigated, but a recent report of very similar evolutionary patterns in plants *A. thaliana* and *Z. mays* [16] highlights the plausibility of such a generalization. Because we have exclusively compared related paralogs to one another and not to orthologous singletons, this observation only describes a fraction of the evolution of duplicates. The concerted expression changes between two members of a paralog pair indeed may or may not be similarly biased towards transcription. Post-duplication expression reduction – likely to restore dosage balance – has been reported [8] and shown to occur at the levels of transcription and translation [9].

We used *in silico* evolution to investigate two potential mechanisms for this predominantly transcriptional divergence of yeast paralogs: an evolutionary trade-off between the cost and precision of gene expression [14] and a larger mutational target size for transcription regulation. Simulations revealed that both hypotheses and their combination may be consistent with the observed patterns of evolution – at least for WGD-derived duplicates. While the minimal model of post-duplication evolution we used was initially described for cases of whole-genome doubling [10], it might also be applicable to other types of duplication events. Our framework for instance partially replicated the divergence patterns of SSD-derived paralog pairs, albeit for distinct parameter combinations and with a lower overall agreement (S10 Fig).

The cost-precision trade-off is sufficient for divergence to occur mostly in transcription, which highlights how interactions at two levels of expression regulation can shape evolutionary trajectories. The general relevance of this type of epistasis in the evolution of duplicate genes is however unclear. Due to the small magnitude of the fitness effects involved [14], cost-precision constraints may only be impactful in large populations where selection is particularly efficient. This might not be the case in the early evolution of WGD-derived paralogs, even in *S. cerevisiae*, as WGD events initially create a much smaller polyploid population. We also note that, contrary to what the N parameter of $10^5 - 10^6$ would suggest, our simulations are not exactly representative of such a large population. Because any beneficial or perfectly neutral mutation is automatically accepted [26], even infinitesimal fitness gains are visible to selection, rather than only those larger than $1/N$. In a more realistic scenario, weakly beneficial and mildly deleterious mutations acting on transcription or translation may both be close to neutrality and have similar fixation probabilities, limiting the ability of the cost-precision trade-off to favor transcriptional divergence.

A larger mutational target size for transcription, modeled as a higher relative mutation probability, emerges as the preferred hypothesis, as it is robust to assumptions about selection efficacy. When strictly considering *cis* mutations – the only changes which can cause two identical gene copies to diverge – as in the current work, a larger mutational target for transcription is intuitively likely. Because *cis* mutations acting on translation have to occur within the transcribed sequence while transcriptional mutations can also arise upstream or downstream of the gene, more nucleotide positions could potentially affect transcription. Direct measurements of the neutral evolutionary rates of transcriptional and translational efficiencies would however be needed to validate this conjecture. Work with yeast hybrids may nonetheless support this idea, as *cis* variants have been shown to significantly affect mRNA abundance more often than translation efficiency [36–39]. This observation nevertheless ought to be considered with caution, especially since translational mutations were often found to have

larger effects on average [37, 39]. In addition, there are even reports that transcription may generally evolve faster across organisms. Differences in translation efficiency have for instance been shown to contribute much less than transcription variation to interspecies expression divergence within primates [40], while almost all expression changes observed over 22 years of bacterial experimental evolution occurred at the transcriptional level [41].

The current distribution of transcription and translation rates among yeast genes has previously been attributed to the optimization of the cost-precision trade-off [14]. Accordingly, suggesting that such constraints are not necessarily involved in the divergence of paralogs and that a difference of mutational target sizes could suffice may appear contradictory. This is especially true considering the significant energetic costs of even small increases of transcription and translation [22, 42]. While any escape of paralog pairs from this latter constraint would need to be thoroughly explained, we note that, even when cost-precision considerations are fully neglected under the minimal model, extended *in silico* evolution results in only minor deviations from the reported distribution of genes in the transcription-translation space (S11 Fig).

Besides providing potential explanations for the expression divergence patterns of yeast paralogs, we also show that the addition of biases to the mutation process – asymmetry in the distribution of mutational effects and correlations between transcriptional and translational effects – may improve their replication by both studied models. This further illustrates how various types of mutational bias could impact the evolutionary trajectories of a duplicate pair under selection to maintain its cumulative protein abundance, as might be the case for singleton genes in the absence of selection [28]. Overall, our work thus highlights the importance of measuring the neutral evolutionary rates of expression at the transcriptional and translational levels, thoroughly characterizing the distributions of mutational effects and assessing between-levels mutational correlations in order to fully understand the expression divergence of paralogs, and by extension, the wider impact of gene duplication on the evolution of gene expression.

Irrespective of its cause(s), the predominantly transcriptional expression divergence we observe within yeast paralogs might have significant evolutionary consequences. Due to the relationship between transcription rate and expression noise [14, 20, 21], it would indeed result in a greater divergence of noise levels within duplicate pairs than symmetrical or mostly translational expression changes. This could consequently favor a specific type of regulatory neofunctionalization which may help explain the observation that 40% of WGD-derived transcription factor pairs in *S. cerevisiae* bind the same targets [43] and further clarify the pivotal role of gene duplication in the growth of eukaryotic regulation networks [44, 45]. This is exemplified by the pair *MSN2*/*MSN4*, which appears to combine the benefits of low and high expression noise. While one gene is stably expressed with low noise across conditions, the other is expressed with high noise and is environmentally responsive [46]. This observation has been interpreted as the resolution of a noise-control conflict in the ancestral function. It is however also conceivable that the benefit of such two-factor regulation was first revealed by the neutral divergence of transcription early after the duplication event and then further refined through changes in promoter architecture. Interestingly, such divergence of stimulus responsiveness between paralogs is not limited to transcription factors – as shown by enzymes *FKS1*/*GSC2* [47] – and may be a general feature of duplicated genes' evolution. Whether this type of evolutionary trajectory is really favored by a bias towards transcription divergence remains to be investigated. Alternatively, the predominance of transcriptional changes we observed could result from widespread selection for such resolution of noise-control conflicts.

We studied the divergence of expression levels between paralogs while neglecting all types of changes to protein function and regulation. This approach would only be suitable if expression changes were not affected by such functional divergence. This would for instance be true if expression diverged first during evolution, in accordance with previous models postulating that early expression changes under selection to maintain cumulative protein abundance lengthen the retention of duplicates and allow function-altering mutations to arise and fix [10, 48]. The absence of a clear continuous relationship between functional divergence and the relative contribution of transcriptional changes may support the plausibility of such models, although stronger relationships with the magnitude of expression

divergence advise caution (S2 Fig). Under this kind of model, expression changes may even shape the subsequent functional divergence of duplicated genes, as the beneficial effect of any mutation affecting the function of a paralog would need to overcome the deleteriousness of the associated reduction in the abundance of the ancestral gene product [10]. Accordingly, such changes may be restricted to new functions for which the protein abundance of the mutated gene copy is already close to optimal. The addition of cost-precision constraints suggests an intriguing extension of this model. If transcription cost and tolerance to expression noise dictate optimal transcription and translation rates for all genes as described by Hausser et al. (2019) [14], functional changes would be further restricted to molecular functions which are compatible with the precision and cost of the affected paralog's expression.

While this work provides insights on the expression divergence of duplicated genes – and thus more generally on the evolution of gene expression –, it presents limitations. A first one is the assumption that all paralog pairs were initially made of two identical gene copies. This is especially significant in the case of WGD-derived pairs, as the yeast WGD likely involved a hybridization event [49]. As such, some paralogs were already diverged and their expression fold-changes may not be representative of the evolution of duplicates. The agreement between WGD- and SSD-derived pairs would however support the idea that a predominance of transcriptional changes is a general feature of the divergence of paralogs. Simplifying assumptions made to simulate the evolution of *S. cerevisiae* duplicated genes also warrant examination. Sequential fixation, where mutants never coexist in the population [24], is likely not an accurate representation of evolutionary processes in yeasts with large population sizes. It is also unable to replicate the fixation dynamics of pairs of compensatory mutations [50], which presumably play an important role in the expression divergence of duplicated genes. More importantly, the use of the Metropolis criterion [26] to accelerate simulations may skew the resulting evolution patterns by equally valuing all positive fitness changes and artificially widening the gap between mildly beneficial and slightly deleterious mutations. Another important limitation of this approach is the absence of gene loss, meaning that many duplicate pairs which would not have been retained over evolutionary timescales contribute to the final simulated divergence distributions. Post-WGD paralog retention is known to be on the order of 15% in *Saccharomyces* yeasts [51], but aiming for both a realistic magnitude of protein abundance divergence and a precise fraction of surviving gene pairs could have made the end condition of the simulations an ever-moving target. Finally, by using the transcription and translation rates reported by Hausser et al. (2019) [14], our work also shares their limitations. Because variations in mRNA and protein degradation were neglected in the corresponding calculations, β_m and β_p are only approximations of the true expression rates. In the case of β_m, since measurements of transcript abundance were used, the reported values – and our assessment of relative divergence – reflect the combined effect of mRNA synthesis and decay. This may be consequential for our test of the cost-precision trade-off, as increases in transcription and decreases in mRNA degradation have distinct effects within this framework [14].

Conclusion

The expression divergence of yeast paralogs mostly occurred at the level of transcription, which may be due to two mechanisms: an evolutionary trade-off between the cost and precision of gene expression and a larger mutational target size for transcription than for translation. Interestingly, the explanatory power of both mechanisms may be improved by the addition of asymmetry and transcription-translation correlations to the distributions of mutational effects, which illustrates the importance of fully characterizing how random mutations affect the different traits contributing to gene expression. Our work moreover highlights how measuring the neutral evolutionary rates of transcription and translation efficiencies is essential to a complete understanding of the evolution of expression levels. Further research will help elucidate whether the greater contribution of transcriptional changes is unique to duplicated genes and clarify its wider implications on gene expression evolution.

Materials and methods

Divergence patterns of duplicated genes

We downloaded the transcription and translation rates of 4440 yeast genes from Hausser et al. (2019) [14]. Paralogs originating from whole-genome duplications (WGD) or small-scale duplication events (SSD) were identified using annotations by Marchant et al. (2019) [52]. Groups of more than two duplicates, derived from successive duplication events, were excluded by only considering genes annotated "WGD" and "SSD" in the cited work. We obtained a final set of 409 high-confidence paralog couples, of which 245 are WGD-derived and 164 have originated from SSD events.

To estimate the relative expression divergence of paralogs, two measures were computed within each pair: log_2 fold-change and normalized difference. Both were calculated for transcription and translation rates as well as for estimated protein abundances.

The log_2 fold-changes were computed according to Eq 1, while the normalized differences were calculated following equation Eq 3, where θ represents any property of interest for genes 1 and 2. The resulting distributions of relative transcriptional and translational divergence were compared in both cases, using the two-sided Mann–Whitney–Wilcoxon test to assess the statistical significance of differences.

$$\Delta_{norm} = \left| \frac{\theta_2 - \theta_1}{\theta_1 + \theta_2} \right| \tag{3}$$

To assess the relationship between the transcriptional and translational levels of divergence, a series of three correlations were computed. In all cases, calculations were done for all duplicates at once as well as separately for WGD- and SSD-derived paralogs.

The first correlation was calculated from the magnitudes of relative divergence. This relative measure was computed for transcription as well as translation rates according to the definitions from Eq 1 and Eq 3 within each paralog pair and Spearman's ρ was calculated from the resulting pairs of transcriptional and translational divergences in both cases.

The second correlation was computed similarly, but this time between signed relative divergences in transcription and translation. Instead of defining ratios as max/min like in Eq 1 and Eq 3, the same genes were assigned arbitrarily to the numerator and denominator for each pair of calculations. This was performed in the two possible orientations for each gene pair and the correlation was computed on the duplicated dataset.

A third and final Spearman's correlation was computed between the log_2 fold-changes of protein abundance (Eq 1) and the divergence ratio D (Eq 2) obtained for each paralog pair. A bootstrapping approach was used to assess the significance of the resulting correlation. Ten thousand shufflings of the dataset were done, during which 409 pairs of paralogs were assembled randomly with replacement from the 818 individual genes. For each newly generated randomized dataset, the calculations of protein abundance log_2 fold-changes, divergence ratios D and the associated correlation were repeated. The p-value was estimated as the fraction of shufflings which resulted in a Spearman's ρ larger or equal to the value computed from the real duplicate pairs. As a validation step, this correlation was also recalculated from signed fold-changes of protein abundance and modified D ratios obtained from signed relative divergences, using a duplicated dataset as previously described.

In addition, a 95% confidence interval was calculated for each of three correlations, using a bootstrapping approach during which 409 true pairs (or 818 when using a duplicated dataset) of observations were sampled with replacement 10 000 times to recompute the corresponding correlation.

Expression and functional divergence

We considered three proxies of functional divergence between paralogs: genetic interaction profiles similarity, overlap of gene ontology (GO) annotation and amino acid sequence divergence.

The matrix of pairwise genetic interaction profile similarities was downloaded (access: 2021-12-17) from TheCellMap.org [53]. Only results from the AllxAll genetic interaction screen were considered. All paralog pairs for which both duplicates were represented in this dataset were kept, leaving 377 gene

pairs. When more than one unique mutant had been screened for one gene, the corresponding vectors of pairwise similarity were averaged. For each paralog pair, we kept the mean of the reported similarity (Pearson correlation) between the interaction profiles of gene P_1 and P_2.

For GO overlap, the GO Slim mappings from the SGD project [54] were downloaded (`http://sgd-archive.yeastgenome.org/curation/literature/`[2022-01-12]). Annotations from the three ontology levels (Process, Function and Component) were combined and the Jaccard index was computed within each paralog pair as the ratio of the intersection over the union of GO terms.

Amino acid sequences for all *S. cerevisiae* ORFs were downloaded from the SGD project (`http://sgd-archive.yeastgenome.org/sequence/S288C_reference/orf_protein/`[2022-01-12]). Pairwise global alignments were performed within each of the 409 paralog pairs using BioPython (v 1.79) and the corresponding amino acid identities were computed.

These three proxies of functional divergence were correlated with the divergence ratios D and log_2 fold-changes of protein abundance calculated previously for each paralog pair using Spearman's ρ.

Minimal model of post-duplication evolution

In agreement with previous work [8, 10], gene duplication is assumed to create two identical gene copies which encode the same protein. Both paralogs thus contribute equally to this protein's cellular function and selection will act to preserve their cumulative protein abundance.

Accordingly with Hausser et al. (2019) [14], an ancestral gene with a transcription rate β_m, in mRNAs per hour, and a translation rate β_p, in proteins per mRNA per hour, is defined. The resulting steady-state protein abundance can be obtained using Eq 4, [14], where α_m and α_p are the mRNA and protein decay rates. They are assumed to be constant and respectively set to 5.10 h^{-1} and 1.34 h^{-1} to account for both active degradation and dilution following cell division [14].

$$p = \frac{\beta_m \beta_p}{\alpha_m \alpha_p} \tag{4}$$

This gene is also described by a unique function $W(p)$ of fitness according to protein abundance. This function is assumed to be a parabola of vertex (p_{opt}, μ), where p_{opt} is an arbitrary optimal protein abundance (in proteins per cell) and μ is the maximal growth rate of 0.42 h^{-1}, whose curvature is described by a noise sensitivity Q defined in Eq 5, [14]. The parameters describing this parabolic fitness function can be calculated as shown in Eq 6.

$$Q = -\frac{1}{2} W''(p_{opt})\, p_{opt} \tag{5}$$

$$\begin{aligned} a &= \frac{W''(p_{opt})}{2} = \frac{-2\left(\frac{Q}{p_{opt}}\right)}{2} = -\frac{Q}{p_{opt}} \\ b &= -2ap_{opt} = -2\left(-\frac{Q}{p_{opt}}\right) p_{opt} = 2Q \\ c &= \mu - ap_{opt}^2 - bp_{opt} = \mu - Qp_{opt} \end{aligned} \tag{6}$$

Following the duplication of any ancestral gene, two paralogs P_1 and P_2 are defined. Each of these copies inherits the ancestral transcription and translation rates, meaning that $\beta_m = \beta_{m1} = \beta_{m2}$ and that $\beta_p = \beta_{p1} = \beta_{p2}$. The total expression is thus doubled. The previously defined parabolic function $W(p)$ becomes the function $W(p_1 + p_2)$ of fitness according to cumulative protein abundance for the duplicate pair. To avoid making any assumptions about the optimality of the duplication, the new protein abundance optimum of this function is set to the smallest product of the ancestral p_{opt} with a constant which ensures $W(p_1 + p_2) > 0$ (see Selecting the post-duplication change in optimal protein abundance). The other parameters of the function (Q and μ) remain unchanged. Accordingly, fitness is computed as:

$$W(p_1 + p_2) = -\frac{Q}{p_{opt}} (p_1 + p_2)^2 + 2Q(p_1 + p_2) + (\mu - Qp_{opt}) \tag{7}$$

Following the duplication event, the expression levels of the two copies evolve under selection to optimize their cumulative protein abundance, as mutations affecting β_m and/or β_p of a single gene at a time reach fixation according to their effect on fitness.

Selecting the post-duplication change in optimal protein abundance

After a gene duplication event, the total transcription of the resulting gene pair is the ancestral transcription rate of the singleton times factor Δ_m. Similarly, the optimal cumulative protein abundance for the two paralogs is Δ_{opt} times the original optimal expression p_{opt} of the ancestral singleton. From Eq 7, a strictly positive post-duplication fitness is thus obtained when:

$$-\frac{Q}{\Delta_{opt}p_{opt}}\left(\Delta_m p_{opt}\right)^2 + 2Q\left(\Delta_m p_{opt}\right) + \mu - Q\Delta_{opt}p_{opt} > 0 \tag{8}$$

Multiplying by Δ_{opt}, an expression of the form $a\Delta_{opt}^2 + b\Delta_{opt} + c$ is obtained:

$$\begin{aligned} &-\frac{Q}{p_{opt}}\left(\Delta_m p_{opt}\right)^2 + 2Q\left(\Delta_m p_{opt}\right)\Delta_{opt} + \mu\Delta_{opt} - Q\Delta_{opt}^2 p_{opt} > 0 \\ &\Rightarrow -Qp_{opt}\Delta_{opt}^2 + \left(2Q\Delta_m p_{opt} + \mu\right)\Delta_{opt} - Q\Delta_m^2 p_{opt} > 0 \end{aligned} \tag{9}$$

To solve for Δ_{opt}, we consider the most extreme case: an ancestral gene with the highest possible noise sensitivity and protein abundance. Accordingly, Q is set to the highest possible value within the framework of Hausser et al. (2019) [14] ($\sim 6.8588 \times 10^{-6}$) and p_{opt} is set to the highest expression level observed in the dataset ($\sim 6.0649 \times 10^6$ proteins per cell). Constant Δ_m is set to 2, meaning a doubling of total transcription, and a maximum growth rate μ of $0.42h^{-1}$ is considered. The following bounds are obtained:

$$\sim 1.87 < \Delta_{opt} < \sim 2.15 \tag{10}$$

In accordance with this result, we used the approximation $\Delta_{opt} = 1.86$ throughout the current work. For all the random seeds used in the simulations, this approximation, obtained for the minimal model, was also valid for the cost-precision model defined below.

Addition of cost-precision constraints

To more fully account for all constraints acting on the expression divergence of duplicated genes, cost-precision constraints were added to the minimal model described above. These implement the trade-off between the cost and the precision of gene expression that was conceptualized by Hausser et al. (2019) [14].

The same assumptions are made as previously, but fitness calculations are complexified to account for transcription cost and stochastic fluctuations of protein abundance.

According to Hausser et al. (2019) [14], the variance of protein abundance for a singleton gene can be approximated by Eq 11, where c_{v0} is a constant representing the minimal coefficient of variation of protein abundance observed in a given species, referred to as noise floor. In *S. cerevisiae*, its value is 0.1.

$$\sigma^2 \approx p^2\left(\frac{1}{p} + \frac{\alpha_p}{\beta_m} + c_{v0}^2\right) \tag{11}$$

For a duplicate pair, the extrinsic and intrinsic components of this expression noise have to be treated separately. While the fluctuations due to the former will be equal for two identical genes in the same cell, those due to the latter will vary independently between the two duplicates. As it does not depend on any gene-specific property, the c_{v0} term is chosen as the extrinsic component (Eq 12).

$$\sigma^2 = \sigma_{int}^2 + \sigma_{ext}^2 \approx p^2\left(\frac{1}{p} + \frac{\alpha_p}{\beta_m}\right) + p^2 c_{v0}^2 \tag{12}$$

Although recent modeling work indicates that this noise floor is extrinsic in nature [55], it might still not fully represent the extrinsic component of expression noise.

The variance on the cumulative protein abundance of P_1 and P_2 is obtained from the variances of their individual protein abundances. In order to perform this calculation, the fluctuations from mean protein abundance across a population of cells can be seen as a random variable with mean 0 and variance σ^2. As shown above (Eq 12), this random variable is itself the sum of two other random variables representing the intrinsic and extrinsic components of these fluctuations. In turn, these two components are each a sum of the respective contributions of both paralogs. For intrinsic noise, the cumulative variance is the sum of the intrinsic variances respectively calculated for each duplicate gene. By definition, intrinsic fluctuations are uncorrelated between duplicates, meaning that the intrinsic components are two independent variables and that their variances can be summed. In contrast, extrinsic fluctuations are the same for two identical paralogs within the same cell, resulting in the extrinsic components of protein abundance variance for P_1 and P_2 being two perfectly positively correlated variables. Their cumulative variance is thus the square of the sum of their standard deviations. Accordingly, the variance of cumulative protein abundance for a duplicate couple is obtained using the following equation:

$$\begin{aligned}\sigma_{tot}^2 &= \sigma_{int1}^2 + \sigma_{int2}^2 + (\sigma_{ext1} + \sigma_{ext2})^2 \\ &\approx p_1^2 \left(\frac{1}{p_1} + \frac{\alpha_p}{\beta_{m1}}\right) + p_2^2 \left(\frac{1}{p_2} + \frac{\alpha_p}{\beta_{m2}}\right) + (p_1 c_{v0} + p_2 c_{v0})^2\end{aligned} \tag{13}$$

For a given protein abundance, the fitness cost of expression varies depending on whether translation is done from few or many mRNAs. Following work by Hausser et al. (2019) [14], the cost of transcription C for two paralogs is calculated from Eq 14, where l_m is the length of the pre-mRNA (identical for both copies), in nucleotides, and c_m is the transcription cost per nucleotide, in nt^{-1}. In the current model, l_m is considered to be a constant and set to the median yeast mRNA length of 1350 nt [14]. The cost per nucleotide c_m is another constant, which has been estimated as $1.2 \times 10^{-9}\ nt^{-1}$ under the assumption that transcriptional resources are limiting and that any increase in transcription level is done at the expense of other transcripts.

$$C = l_m c_m \left(\beta_{m1} + \beta_{m2}\right) \tag{14}$$

Taking expression noise and transcription costs into account, the fitness $F\left(P_{tot}\right)$ of a population of cells at mean cumulative protein abundance P_{tot} becomes the mean of fitness W under protein abundance variance σ_{tot}^2 minus the penalty C:

$$F\left(P_{tot}\right) = E\left(W\left(p_{tot}\right)\right) - C \tag{15}$$

Since the variance σ_{tot}^2 of cumulative protein abundance p_{tot} and the function linking this variable to fitness are known, the mean (populational) fitness can be computed. This mean fitness $E\left(W\left(p_{tot}\right)\right)$ can be expressed as Eq16, where a, b and c are the parameters of the parabolic fitness function $W\left(p_{tot}\right)$. The mean of the squared cumulative protein abundance $E\left(p_{tot}^2\right)$ is obtained according to Eq 17, using the definition of variance as the difference between the mean of the squares and the square of the mean. The populational fitness at a given mean cumulative abundance can thus be calculated according to Eq 18.

$$E\left(W\left(p_{tot}\right)\right) = aE\left(p_{tot}^2\right) + bE\left(p_{tot}\right) + c \tag{16}$$

$$E\left(p_{tot}^2\right) = \sigma_{tot}^2 + P_{tot}^2 \tag{17}$$

$$E\left(W\left(p_{tot}\right)\right) = a\left(\sigma_{tot}^2 + P_{tot}^2\right) + bP_{tot} + c \tag{18}$$

Framework to simulate post-duplication regulatory evolution

To simulate the expression divergence of yeast paralogs, a sequential fixation approach is used [24]. We thus make the simplifying assumption that the mutation rate is low enough for only two alleles (the

ancestral state and a mutant) of a given duplicate pair to coexist simultaneously in the population. Successive mutation rounds are done, and on each occasion, transcription and/or translation rates of one paralog are mutated randomly. This new expression level either reaches fixation or is lost before another mutation is attempted, according to probabilities calculated from its effect on fitness.

Initialization

In order to obtain an arbitrarily small or large number of paralog pairs, the same number n of ancestral singleton genes must first be generated. This is done through the initialization of an array of n rows containing ancestral β_m and β_p values. Thus, no nucleotide or amino acid sequences are modeled – only their expression phenotypes.

Distinct groups of ancestral genes are generated depending on whether their evolution is simulated according to the minimal model or to its cost-precision implementation, but all are obtained from the same combinations of ancestral expression optima p_{opt} and noise sensitivities Q.

Combinations of protein abundance p_{opt} – obtained from the reported transcription and translation rates – and Q are first sampled n times with replacement from the 4440 individual genes (singletons as well as duplicates) included in the dataset. Values of Q had previously been inferred for all genes using Eq 19, following work by Hausser et al. (2019) [14]. For this sampling, values of Q above the theoretical maximum reported by Hausser et al. (2019) [14] ($\sim 6.8588 \times 10^{-6}$) are excluded. Combinations of p_{opt} and noise sensitivity resulting in a fitness function curvature below a specified threshold are also filtered out, to avoid cases where a two-fold reduction of cumulative expression immediately after the duplication would be neutral or beneficial (see Selection of valid ancestral genes).

$$Q = c_m \alpha_m l_m \frac{\beta_m}{\beta_p} \tag{19}$$

To obtain ancestral singletons for simulations under the minimal model of post-duplication evolution, transcription and translation rates are computed following Eq 19. It is used to calculate an optimal $\frac{\beta_p}{\beta_m}$ ratio according to the evolutionary trade-off formalised by Hausser et al. (2019) [14]. The transcription rate β_m and translation rate β_p which satisfy both this ratio and the optimal expression p_{opt} are then computed. Strictly speaking, an infinity of combinations of β_m and β_p are optimal for any gene under the minimal model, as long as they result in the specified protein abundance p_{opt}. We thus use Eq 19 to reproducibly choose one realistic combination for each ancestral singleton. Because the minimal model should not account in any way for the cost-precision constraints on gene expression, Q values are subsequently resampled with replacement from the genomic distribution. As previously, values above the theoretical maximum are excluded and the sampling is repeated if the new Q results in an invalid combination of noise sensitivity and p_{opt}.

To create ancestral genes for simulations under the cost-precision trade-off, the first combinations of Q and p_{opt} generated are again taken as a starting point. This time, corresponding gene-specific functions $W(p)$ of fitness according to expression level (Eq 15) are used to identify the β_m and β_p combination which maximizes fitness. This optimization is performed using a differential evolution algorithm – the *differential_evolution* method of the *optimize* suite of the *SciPy* module [56]. Although Eq 19 already describes an optimal β_m-β_p pair, the latter may not correspond to the true optimum, as it is restricted to values which result in protein abundance p_{opt}. Since the magnitude of stochastic expression fluctuations (noise) scales with protein abundance (Eq 11), the true optimal expression under the cost-precision trade-off may indeed be slightly below the abundance optimum p_{opt}.

Once generated, all ancestral genes are duplicated as previously described into two paralogs P_1 and P_2, both with the ancestral rates β_m and β_p. For each duplicate pair, a new function $W(p_{tot})$ of fitness according to cumulative protein abundance is defined and its optimum is set to the ancestral p_{opt} times $\Delta_{opt} = 1.86$.

Selection of valid ancestral genes

During the generation of ancestral singletons, a minimal threshold of fitness function curvature is enforced. This ensures that all selected genes are sensitive enough to changes in protein abundance for the immediate post-duplication loss of a paralog to be deleterious. A duplicate pair for which this would not be the case would rapidly revert to the singleton state.

Classical population genetics theory indicates that a mutation needs to cause a loss of fitness greater than the inverse of the effective population size to be efficiently selected against. Accordingly, we want to identify conditions under which the loss of a paralog immediately after duplication would reduce fitness by more than $1/N$. Under the minimal model, that is:

$$W\left(p_{tot}\right) - W\left(\frac{p_{tot}}{2}\right) > \frac{1}{N} \tag{20}$$

For the filtering of singleton genes, it is more convenient to express p_{tot} the product of ancestral protein abundance optimum p_{opt} and constant Δ_m, which describes the post-duplication change of total transcription and thereby of total protein. Using the parabola described in Eq 6 as the fitness function $W\left(p_{opt}\right)$ and also adding constant Δ_{opt} – describing the post-duplication change of the optimal cumulative protein abundance – to generalize to any duplication, we obtain:

$$\left(-\frac{Q}{\Delta_{opt}p_{opt}}\left(\Delta_m p_{opt}\right)^2 + 2Q\left(\Delta_m p_{opt}\right)\right) - \left(-\frac{Q}{\Delta_{opt}p_{opt}}\left(\frac{\Delta_m p_{opt}}{2}\right)^2 + 2Q\left(\frac{\Delta_m p_{opt}}{2}\right)\right) > \frac{1}{N} \tag{21}$$

Summing and simplifying, we obtain the following expression:

$$\Delta_m Q p_{opt}\left(1 - \frac{3\Delta_m}{4\Delta_{opt}}\right) > \frac{1}{N} \tag{22}$$

Accordingly, all ancestral singletons included in the current simulations combine a noise sensitivity Q and a protein abundance optimum p_{opt} which satisfy the following condition:

$$Q p_{opt} > \frac{1}{c_n N} \tag{23}$$

$$where\ c_n = \Delta_m\left(1 - \frac{3\Delta_m}{4\Delta_{opt}}\right)$$

This condition is only valid when $c_n > 0$, which implies that $\frac{3\Delta_m}{4\Delta_{opt}} < 1$. In accordance with this, all simulations presented in the current work are done under $\frac{3}{4}\Delta_m < \Delta_{opt}$.

Mutation-selection approach

Sequential fixation simulations are made on the two sets of duplicate pairs generated, through a succession of mutation-selection rounds. One random mutation is first sampled for each n paralog pair from a normal distribution with mean 0 and arbitrary standard deviation σ_{mut}. These mutational effects are each assigned randomly to transcription or translation rates according to relative probabilities $P_{\beta m}$ and $P_{\beta p}$, which represent the relative mutational target sizes of the two traits. For instance, $P_{\beta m} = 2$ and $P_{\beta p} = 1$ would be used in a simulation where the mutational target size of transcription is assumed to be twice that of translation. Once it has been defined as either transcriptional or translational, each mutation is assigned at random to one of paralogs P_1 and P_2 of the corresponding duplicate pair, which are both equally likely to be mutated. We note that these steps of mutation generation and assignment are done at once for the two sets of simulated gene pairs. As such, the nth paralog couple of both simulations – respectively made according to the minimal and cost-precision models – receives the same precise series of mutations.

So that their impact on fitness can be assessed, the sampled mutational effects are then applied to the transcription or translation rates of the designated gene copies across the minimal model and

cost-precision simulations. Epistasy is assumed to be multiplicative, so that all mutations have relative effects and are applied as shown in Eq 24, where δ_m and δ_p respectively represent the transcriptional and translational magnitudes of the mutation. Because a given mutation affects β_m or β_p but not both simultaneously, at least one of δ_m and δ_p will be null every time a mutational effect is applied.

$$\begin{aligned} \beta_{mj} &= \beta_{mi} + \delta_m \beta_{mi} \\ \beta_{pj} &= \beta_{pi} + \delta_p \beta_{pi} \end{aligned} \tag{24}$$

These new transcription and translation rates are used to compute mutant fitness values F_j for all paralog pairs of the two simulations, accordingly with the specifications of the minimal and cost-precision models. Thus, while the same mutations are attempted across the two models, they do not necessarily have the same fitness effects in both scenarios. Mutant fitness F_j is then compared to ancestral fitness values F_i computed using the pre-mutation β_m and β_p of all duplicates and a fixation probability P_{fix} is calculated for each mutation. This is done using the Metropolis criterion to accelerate the simulation [26]. Following Eq 25, any beneficial or completely neutral mutation is automatically accepted, while deleterious mutations are fixed with a probability which decreases exponentially according to the magnitude of their fitness effect.

$$P_{fix} = \begin{cases} 1 & \text{if } F_j > F_i \\ e^{-2N(log(F_i) - log(F_j))} & \text{if } F_j \leq F_i \end{cases} \tag{25}$$

Prior to this fixation probability calculation, all fitness values, which are growth rates between 0 and 0.42 h^{-1}, are scaled between 0 and 1. The same set of randomly generated floats is used in every simulation to decide whether each mutation is rejected or reaches fixation according to P_{fix}. In addition, any mutation taking β_m or β_p to or below 0 or increasing any of the two rates above the maximal value found in the *S. cerevisiae* dataset [14] is automatically rejected. Once the fate of each mutation has been established, the new transcription and translation rates of all simulated paralogs are set and this process of mutation-selection is repeated.

Simulations are stopped as soon as they have resulted in a realistic magnitude of protein abundance divergence. Following each mutation-selection round, the log_2 fold-change of protein abundance is calculated for every simulated duplicate pair using Eq 1. Mood's median test is used to compare this distribution to the empirical protein expression divergence previously computed in the same manner, either for WGD- or SSD-derived paralogs or for all pairs at once. Once a p-value ≥ 0.1 is obtained, the simulated protein abundance divergence is considered to have reached a realistic magnitude and the simulation is stopped. This is done separately for the two models, so that their respective simulations may not be completed in the same number of rounds.

With asymmetry in the distribution of mutational effects

For mutations increasing or decreasing expression to not necessarily be equally likely, the mutation-selection framework described above was slightly modified. The normal distribution from which mutational effects are sampled was replaced by a skew normal distribution of mean 0, standard deviation σ_{mut} and skewness parameter $\alpha \neq 0$. No changes were made to the steps of mutation assignment and selection.

With correlated mutational effects

To implement arbitrary correlations between the transcriptional and translational effects of mutations, the sampling of mutational effects was performed on a bivariate normal distribution of means 0 and standard deviations $\sigma_{\beta m}$ and $\sigma_{\beta p}$. Because mutations now affected transcription and translation simultaneously, the step of assigning each to one level of regulation of the other with relative probabilities $P_{\beta m}$ and $P_{\beta p}$ had to be skipped. Consequently, the way differences in mutational target sizes were modeled was modified. Instead of using the probabilities of transcriptional and translational

mutations, the magnitudes of their effects was utilized. A larger mutational target size was thus implemented as a higher corresponding standard deviation of mutational effects. For instance, a target size twice larger for transcription than for translation was modeled as $\sigma_{\beta m} = 2\sigma_{\beta p}$.

The precise values of $\sigma_{\beta m}$ and $\sigma_{\beta p}$ were set to ensure a roughly constant expected protein abundance change per mutation across all simulations. To this end, a brute-force optimization approach – the *brute* method of the *optimize* suite of the *SciPy* module [56] – was used to find the $\sigma_{\beta m}$ and $\sigma_{\beta p}$ values which verified the desired mutational target size ratio while resulting in the same mean absolute protein abundance change per mutation as a chosen reference σ_{mut} in the general framework. The covariance matrix of the bivariate normal distribution was computed from the two standard deviations resulting from this optimization, according to a specified correlation coefficient r_{mut} between δ_m and δ_p.

Assessment of the simulations

Once simulations are completed, log_2 fold-changes (Eq 1) of transcription rate, translation rate and protein abundance as well as divergence ratio D (Eq 2) are computed for all simulated gene pairs. The resulting distributions are compared to their empirical counterparts – obtained for the true paralog couples (either WGD, SSD or all) – in two ways. The KS statistic is computed to quantify the fit between the distributions, while Mood's median test is performed to assess whether the median of the simulated distribution is of a realistic magnitude.

The log_2 fold-changes – including their signed version – and divergence ratios obtained for the simulated duplicates are also used to compute the three correlations described previously. These are compared to the corresponding 95% confidence interval obtained for the true gene pairs, either when considering only WGD- or SSD-derived paralogs or for all duplicates. Because the two paralogs in a simulated pair are ordered randomly, signed log_2 fold-changes are computed in only one arbitrary orientation for each replicate simulation.

Simulations

Test of the cost-precision trade-off

A small-scale simulation was performed for 50 randomly generated paralog pairs according to both models. High selection efficacy ($N = 1 \times 10^6$) and small mutational effects ($\sigma_{mut} = 0.025$) were postulated, and mutations affecting transcription or translation were assumed to be equally likely ($P_{\beta m} = P_{\beta p}$). The end condition was evaluated according to the divergence of WGD-derived *S. cerevisiae* paralog pairs.

The transcription and translation rates of each simulated gene as well as the associated fitness values were kept for each mutation-selection round. Evolutionary trajectories for both expression rates were obtained by the normalization of each variable by its initial value, followed by a log_2 transformation. Fitness values were min-max normalized individually for each simulated paralog pair.

Identification of the best parameters and comparison of the models

To simultaneously identify the best-fitting parameters and compare the performance of both models, a grid-search of σ_{mut} and mutational target size ratios $P_{\beta m}/P_{\beta p}$ was performed. We tested values {0.010, 0.025, 0.050, 0.075, 0.100, 0.125, 0.150, 0.200, 0.250, 0.350} for σ_{mut} and {1/2, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10} for $P_{\beta m}/P_{\beta p}$. For each combination of parameters, three replicate simulations of 2000 paralog pairs were made, according to both the minimal and cost-precision models. The same set of three random seeds was used for all trios of simulations, such that all differences between runs are due to changes in parameter values. This grid-search process was done twice, under the assumption of high ($N = 1 \times 10^6$) and reduced ($N = 1 \times 10^5$) selection efficacy. Because we focused on WGD-derived paralogs, the end condition for each of these simulations was set according to the relative protein abundance divergence of this category of duplicates.

The most realistic – or rather best-fitting – σ_{mut} was identified individually for both values of N by comparing the simulation results to the empirical divergence patterns of *S. cerevisiae* WGD-derived paralogs. This best σ_{mut} was defined as the value which, for any of the two models, resulted in the lowest overall grand mean KS statistic (mean KS value for β_m, β_p protein abundance and D across three replicate simulations) for any $P_{\beta m}/P_{\beta p}$ while also generating simulated relative divergence distributions whose medians were not significantly different from the empirical ones (p-val > 0.05, Mood's median test) for β_m, β_p and D for every replicate simulation. Protein abundance was not considered, as the end condition already guaranteed the realism of its magnitude in every case.

Once the two best-fitting σ_{mut}, respectively for $N = 1 \times 10^6$ and $N = 1 \times 10^5$, had been identified, the corresponding simulation results were used to compare the performance of both models according to their replication of the four features of the empirical divergence of paralogs, using WGD-derived ones as a comparative.

Additional simulations at $N = 1 \times 10^6$ were performed using the associated best σ_{mut} across an extended range of $P_{\beta m}/P_{\beta p}$ favoring translational mutations: $\{1/10, 1/6, 1/5, 1/4, 1/3, 1/2\ 1, 2, 3, 4\}$. These simulations were also done in three replicates of 2000 paralogs and assessed according to WGD-derived duplicates.

Finally, identical grid-searches of σ_{mut} and $P_{\beta m}/P_{\beta p}$ were also performed for SSD-derived paralogs. In this case, the stopping of replicate simulations as well as the identification of the best σ_{mut} and the final measurement of model performance were done according to the SSD-derived duplicates of *S. cerevisiae*.

Impact of asymmetrically distributed mutational effects

To assess the impact of asymmetry in the distribution of mutational effects, further simulations were done using the modified mutation-selection approach which samples mutations from a skew normal distribution. A range of negative values $\{-0.010, -0.050, -0.075, -0.100, -0.125, -0.150, -0.175, -0.250, -0.350\}$ were considered for the skewness parameter α. These simulations were performed in grid along a range of mutational target size ratios $P_{\beta m}/P_{\beta p}$ $\{1/2, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ at the best-fitting σ_{mut} identified for WGD-derived paralogs. For each combination of parameters, three replicate simulations of 2000 randomly generated paralog pairs were done as previously described under both scenarios of selection efficacy. The end condition of these simulations was set according to WGD-derived duplicates, and they were also evaluated by comparison to this group of gene pairs.

Impact of correlations between transcriptional and translational mutations

To test the effect of various levels of correlation between the transcriptional and translational effects of mutations, an additional grid-search of evolutionary parameters first had to be performed using the mutation-selection framework adapted to use a bivariate distribution of mutational effects. We considered a range of reference σ_{mut} $\{0.010, 0.025, 0.050, 0.075, 0.100, 0.125, 0.150, 0.200, 0.250, 0.350\}$ according to which the standard deviations $\sigma_{\beta m}$ and $\sigma_{\beta p}$ were set. As previously, $P_{\beta m}/P_{\beta p}$ mutational target size ratios $\{1/2, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ were tested. Three replicate simulations of 2000 duplicate pairs were performed for each combination of parameters at both selection efficacies N and were assessed according to the expression divergence of WGD-derived paralogs. The best-fitting reference σ_{mut} was identified using the same definition as before.

Additional series of grid simulations were performed – one for each N – using the best reference σ_{mut} respectively identified for the two levels of selection efficacy. The same range of $P_{\beta m}/P_{\beta p}$ ratios was combined with a range of correlation coefficients r_{mut} between the transcriptional and translational effects of mutations: $\{-0.4, -0.3, -0.2, -0.1, 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$. These simulations were also performed in three replicates of 2000 duplicate pairs as described previously and evaluated by comparison to the WGD-derived paralogs of *S. cerevisiae*.

Code availability

All data preparation, simulations and analysis were done in Python (v 3.8.5), and all statistical tests were performed using their *SciPy* (v 1.6.2) implementation. The corresponding code, as well as all the batch files used to call the simulation script, have been deposited on GitHub: `https://github.com/Landrylab/Aube_et_al_2022`.

Supporting information

S1 Fig. Validating the expression divergence patterns of yeast paralogs.

S2 Fig. The predominance of transcriptional divergence is only weakly associated with the level of functional divergence among *S. cerevisiae* duplicated genes.

S3 Fig. Identification of the best-fitting σ_{mut} across models and selection efficacies.

S4 Fig. Replication of relative divergence in transcription, translation and protein abundance as well as of the divergence ratio across models and levels of selection efficacy for the corresponding best-fitting σ_{mut} values.

S5 Fig. A larger mutational target size for translation does not rescue translational divergence under the cost-precision model when the efficacy of selection is high.

S6 Fig. The postulated post-duplication change of optimal cumulative protein abundance has limited influence on the divergence correlations generated from simulations.

S7 Fig. Identification of the best-fitting reference σ_{mut} across models and selection efficacies when bivariate mutational effects are used.

S8 Fig. The addition of realistic mutational constraints also allows the minimal model to replicate four features of the empirical divergence of yeast WGD-derived paralogs under conditions of reduced selection efficacy.

S9 Fig. The addition of realistic mutational constraints allows the cost-precision trade-off to simultaneously replicate four features of the empirical divergence patterns of yeast paralogs, but only when the efficacy of selection is reduced ($N = 1 \times 10^5$).

S10 Fig. The minimal and cost-precision models offer some replication of the divergence patterns of SSD-derived yeast paralogs, but for distinct parameter combinations and with a lower overall agreement.

S11 Fig. Simulations under both models generate realistic distributions of genes in the expression space.

S1 File. Animation of the divergence of transcription and translation under the minimal model of post-duplication evolution. Mutations cause small changes in the transcription and translation rates of paralogs P_1 and P_2. Selection to maintain cumulative expression however maintains their total protein abundance at an approximately constant level.

Acknowledgments

We thank all Landry lab members, and especially Angel Cisneros, Philippe Després and Johan Hallin, for insightful journal club discussions which led to this project. Funding for this work was from a NSERC discovery grant to CRL. SA was supported by graduate scholarships from NSERC and FRQNT, as well as a PBLDD scholarship from Université Laval. LNT was supported by an Alexander Graham Bell doctoral scholarship from NSERC. CRL holds the Canada Research Chair in Cellular Systems and Synthetic Biology.

References

1. Deutschbauer AM, Jaramillo DF, Proctor M, Kumm J, Hillenmeyer ME, Davis RW, et al. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. Genetics. 2005;169(4):1915–1925. doi:10.1534/genetics.104.036871.

2. Huang N, Lee I, Marcotte EM, Hurles ME. Characterising and Predicting Haploinsufficiency in the Human Genome. PLOS Genetics. 2010;6(10):e1001154. doi:10.1371/journal.pgen.1001154.

3. Keren L, Hausser J, Lotan-Pompan M, Vainberg Slutskin I, Alisar H, Kaminski S, et al. Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. Cell. 2016;166(5):1282–1294.e18. doi:10.1016/j.cell.2016.07.024.

4. Morrill SA, Amon A. Why haploinsufficiency persists. Proceedings of the National Academy of Sciences. 2019;116(24):11866–11871. doi:10.1073/pnas.1900437116.

5. Hill MS, Vande Zande P, Wittkopp PJ. Molecular and evolutionary processes generating variation in gene expression. Nature Reviews Genetics. 2021;22(4):203–215. doi:10.1038/s41576-020-00304-w.

6. Ohno S. Evolution by gene duplication. London: G. Allen and Unwin ;; 1970.

7. Force A, Lynch M, Pickett FB, Amores A, Yan Yl, Postlethwait J. Preservation of Duplicate Genes by Complementary, Degenerative Mutations. Genetics. 1999;151(4):1531–1545.

8. Qian W, Liao BY, Chang AYF, Zhang J. Maintenance of duplicate genes and their functional redundancy by reduced expression. Trends in genetics : TIG. 2010;26(10):425–430. doi:10.1016/j.tig.2010.07.002.

9. Chang AYF, Liao BY. Reduced Translational Efficiency of Eukaryotic Genes after Duplication Events. Molecular Biology and Evolution. 2020;37(5):1452–1461. doi:10.1093/molbev/msz309.

10. Gout JF, Lynch M. Maintenance and Loss of Duplicated Genes by Dosage Subfunctionalization. Molecular Biology and Evolution. 2015;32(8):2141–2148. doi:10.1093/molbev/msv095.

11. Thompson A, Zakon HH, Kirkpatrick M. Compensatory Drift and the Evolutionary Dynamics of Dosage-Sensitive Duplicate Genes. Genetics. 2016;202(2):765–774. doi:10.1534/genetics.115.178137.

12. Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, et al. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proceedings of the National Academy of Sciences. 2008;105(27):9272–9277. doi:10.1073/pnas.0803466105.

13. Lipinski KJ, Farslow JC, Fitzpatrick KA, Lynch M, Katju V, Bergthorsson U. High Spontaneous Rate of Gene Duplication in Caenorhabditis elegans. Current biology : CB. 2011;21(4):306–310. doi:10.1016/j.cub.2011.01.026.

14. Hausser J, Mayo A, Keren L, Alon U. Central dogma rates and the trade-off between precision and economy in gene expression. Nature Communications. 2019;10(1):68. doi:10.1038/s41467-018-07391-8.

15. Wolfe KH, Shields DC. Molecular evidence for an ancient duplication of the entire yeast genome. Nature. 1997;387(6634):708–713. doi:10.1038/42711.

16. Wang S, Chen Y. Fine-Tuning the Expression of Duplicate Genes by Translational Regulation in Arabidopsis and Maize. Frontiers in Plant Science. 2019;10:534. doi:10.3389/fpls.2019.00534.

17. Hough J, Williamson RJ, Wright SI. Patterns of Selection in Plant Genomes. Annual Review of Ecology, Evolution, and Systematics. 2013;44(1):31–49. doi:10.1146/annurev-ecolsys-110512-135851.

18. Tsai IJ, Bensasson D, Burt A, Koufopanou V. Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle. Proceedings of the National Academy of Sciences. 2008;105(12):4957–4962. doi:10.1073/pnas.0707314105.

19. Skelly DA, Ronald J, Connelly CF, Akey JM. Population Genomics of Intron Splicing in 38 Saccharomyces cerevisiae Genome Sequences. Genome Biology and Evolution. 2009;1:466–478. doi:10.1093/gbe/evp046.

20. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. Regulation of noise in the expression of a single gene. Nature Genetics. 2002;31(1):69–73. doi:10.1038/ng869.

21. Blake WJ, Kaern M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. Nature. 2003;422(6932):633–637. doi:10.1038/nature01546.

22. Lynch M, Marinov GK. The bioenergetic costs of a gene. Proceedings of the National Academy of Sciences. 2015;112(51):15690–15695. doi:10.1073/pnas.1514974112.

23. Kafri M, Metzl-Raz E, Jona G, Barkai N. The Cost of Protein Production. Cell Reports. 2016;14(1):22–31. doi:10.1016/j.celrep.2015.12.015.

24. McCandlish DM, Stoltzfus A. Modeling evolution using the probability of fixation: history and implications. The Quarterly Review of Biology. 2014;89(3):225–252. doi:10.1086/677571.

25. Taggart JC, Li GW. Production of Protein-Complex Components Is Stoichiometric and Lacks General Feedback Regulation in Eukaryotes. Cell Systems. 2018;7(6):580–589.e4. doi:10.1016/j.cels.2018.11.003.

26. Teufel AI, Wilke CO. Accelerated simulation of evolutionary trajectories in origin-fixation models. Journal of The Royal Society Interface. 2017;14(127):20160906. doi:10.1098/rsif.2016.0906.

27. Metzger BPH, Duveau F, Yuan DC, Tryban S, Yang B, Wittkopp PJ. Contrasting Frequencies and Effects of cis- and trans-Regulatory Mutations Affecting Gene Expression. Molecular Biology and Evolution. 2016;33(5):1131–1146. doi:10.1093/molbev/msw011.

28. Hodgins-Davis A, Duveau F, Walker EA, Wittkopp PJ. Empirical measures of mutational effects define neutral models of regulatory evolution in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences. 2019;116(42):21085–21093. doi:10.1073/pnas.1902823116.

29. Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL. Genetic properties influencing the evolvability of gene expression. Science (New York, NY). 2007;317(5834):118–121. doi:10.1126/science.1140247.

30. Vaishnav ED, de Boer CG, Molinet J, Yassour M, Fan L, Adiconis X, et al. The evolution, evolvability and engineering of gene regulatory DNA. Nature. 2022;603(7901):455–463. doi:10.1038/s41586-022-04506-6.

31. Chen S, Li K, Cao W, Wang J, Zhao T, Huan Q, et al. Codon-Resolution Analysis Reveals a Direct and Context-Dependent Impact of Individual Synonymous Mutations on mRNA Level. Molecular Biology and Evolution. 2017;34(11):2944–2958. doi:10.1093/molbev/msx229.

32. Chan LY, Mugler CF, Heinrich S, Vallotton P, Weis K. Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. eLife. 2018;7:e32536. doi:10.7554/eLife.32536.

33. Preiss T, Baron-Benhamou J, Ansorge W, Hentze MW. Homodirectional changes in transcriptome composition and mRNA translation induced by rapamycin and heat shock. Nature Structural & Molecular Biology. 2003;10(12):1039–1047. doi:10.1038/nsb1015.

34. Halbeisen RE, Gerber AP. Stress-Dependent Coordination of Transcriptome and Translatome in Yeast. PLOS Biology. 2009;7(5):e1000105. doi:10.1371/journal.pbio.1000105.

35. Blevins WR, Tavella T, Moro SG, Blasco-Moreno B, Closa-Mosquera A, Díez J, et al. Extensive post-transcriptional buffering of gene expression in the response to severe oxidative stress in baker's yeast. Scientific Reports. 2019;9:11005. doi:10.1038/s41598-019-47424-w.

36. Albert FW, Muzzey D, Weissman JS, Kruglyak L. Genetic Influences on Translation in Yeast. PLOS Genetics. 2014;10(10):e1004692. doi:10.1371/journal.pgen.1004692.

37. Artieri CG, Fraser HB. Evolution at two levels of gene expression in yeast. Genome Research. 2014;24(3):411–421. doi:10.1101/gr.165522.113.

38. McManus CJ, May GE, Spealman P, Shteyman A. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Research. 2014;24(3):422–430. doi:10.1101/gr.164996.113.

39. Wang Z, Sun X, Zhao Y, Guo X, Jiang H, Li H, et al. Evolution of gene regulation during transcription and translation. Genome Biology and Evolution. 2015;7(4):1155–1167. doi:10.1093/gbe/evv059.

40. Wang SH, Hsiao CJ, Khan Z, Pritchard JK. Post-translational buffering leads to convergent protein expression levels between primates. Genome Biology. 2018;19(1):83. doi:10.1186/s13059-018-1451-z.

41. Favate JS, Liang S, Yadavalli SS, Shah P. The landscape of transcriptional and translational changes over 22 years of bacterial adaptation. bioRxiv. 2021; p. 2021.01.12.426406. doi:10.1101/2021.01.12.426406.

42. Wagner A. Energy constraints on the evolution of gene expression. Molecular Biology and Evolution. 2005;22(6):1365–1374. doi:10.1093/molbev/msi126.

43. Gera T, Jonas F, More R, Barkai N. Evolution of binding preferences among whole-genome duplicated transcription factors. eLife. 2022;11:e73225. doi:10.7554/eLife.73225.

44. Teichmann SA, Babu MM. Gene regulatory network growth by duplication. Nature Genetics. 2004;36(5):492–496. doi:10.1038/ng1340.

45. Charoensawan V, Wilson D, Teichmann SA. Genomic repertoires of DNA-binding transcription factors across the tree of life. Nucleic Acids Research. 2010;38(21):7364–7377. doi:10.1093/nar/gkq617.

46. Chapal M, Mintzer S, Brodsky S, Carmi M, Barkai N. Resolving noise–control conflict by gene duplication. PLOS Biology. 2019;17(11):e3000289. doi:10.1371/journal.pbio.3000289.

47. Mazur P, Morin N, Baginsky W, el Sherbeini M, Clemas JA, Nielsen JB, et al. Differential expression and function of two homologous subunits of yeast 1,3-beta-D-glucan synthase. Molecular and Cellular Biology. 1995;15(10):5671–5681. doi:10.1128/MCB.15.10.5671.

48. Teufel AI, Liu L, Liberles DA. Models for gene duplication when dosage balance works as a transition state to subsequent neo- or sub-functionalization. BMC Evolutionary Biology. 2016;16(1):45. doi:10.1186/s12862-016-0616-1.

49. Marcet-Houben M, Gabaldón T. Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. PLOS Biology. 2015;13(8):e1002220. doi:10.1371/journal.pbio.1002220.

50. Carter AJR, Wagner GP. Evolution of functionally conserved enhancers can be accelerated in large populations: a population-genetic model. Proceedings Biological Sciences. 2002;269(1494):953–960. doi:10.1098/rspb.2002.1968.

51. Scannell DR, Frank AC, Conant GC, Byrne KP, Woolfit M, Wolfe KH. Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(20):8397–8402. doi:10.1073/pnas.0608218104.

52. Marchant A, Cisneros AF, Dubé AK, Gagnon-Arsenault I, Ascencio D, Jain H, et al. The role of structural pleiotropy and regulatory evolution in the retention of heteromers of paralogs. eLife. 2019;8:e46754. doi:10.7554/eLife.46754.

53. Usaj M, Tan Y, Wang W, VanderSluis B, Zou A, Myers CL, et al. TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the Global Yeast Genetic Interaction Network. G3 Genes|Genomes|Genetics. 2017;7(5):1539–1549. doi:10.1534/g3.117.040220.

54. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, et al. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Research. 2012;40(Database issue):D700–705. doi:10.1093/nar/gkr1029.

55. Jedrak J, Ochab-Marcinek A. Contributions to the 'noise floor' in gene expression in a population of dividing cells. Scientific Reports. 2020;10(1):13533. doi:10.1038/s41598-020-69217-2.

56. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods. 2020;17(3):261–272. doi:10.1038/s41592-019-0686-2.

Supplementary figures

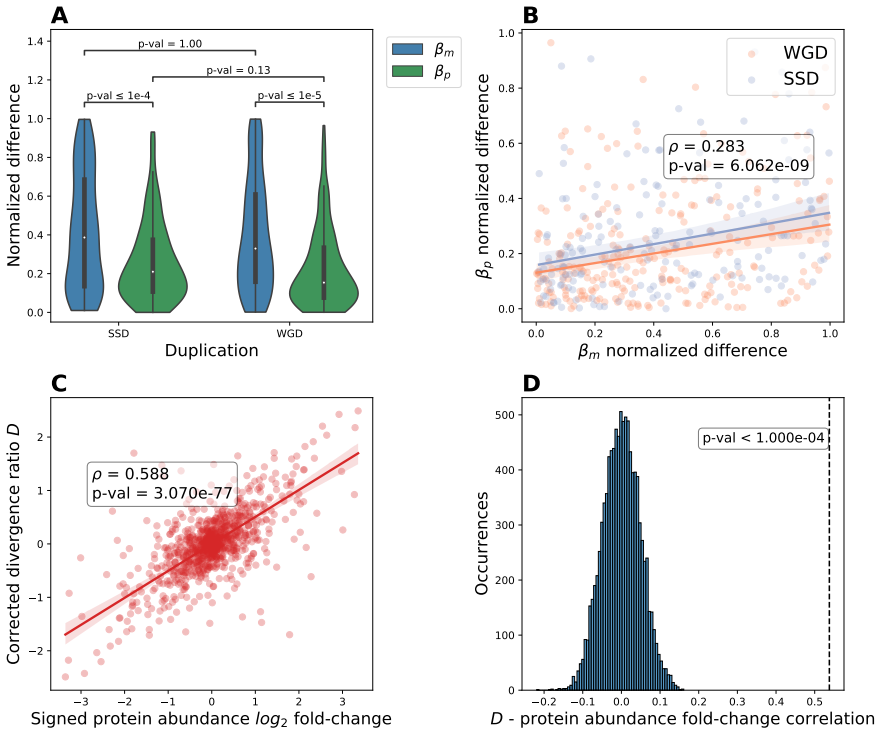

Fig S1. Validating the expression divergence patterns of yeast paralogs. (A) Distributions of within-couple relative divergence for *S. cerevisiae* paralogs according to the associated mechanism of duplication (245 WGD pairs and 164 SSD ones), when computed as normalized differences. Statistical significance of difference assessed with Mann-Whitney-Wilcoxon two-sided tests. (B) Correlation between the magnitude of transcriptional and translational relative divergence for WGD- and SSD-derived yeast paralog pairs, when computed from normalized differences. Spearman's ρ and p-value shown for all duplicate pairs. (C) Correlation between the divergence ratio and the protein abundance log_2 fold-change for all paralog pairs when both values are computed in a way that allows negative values. Calculations were made in the two possible orientations for each gene pair and Spearman's ρ and p-value were computed from the resulting duplicated dataset (n=818). (D) Distributions of correlations between the divergence ratio and the log_2 fold-change of protein abundance (Fig 1C) obtained from 10 000 randomizations of 409 paralog pairs from the complete set of *S. cerevisiae* duplicated genes. The dashed line represents the value obtained for the 409 true gene couples and the corresponding p-value is shown.

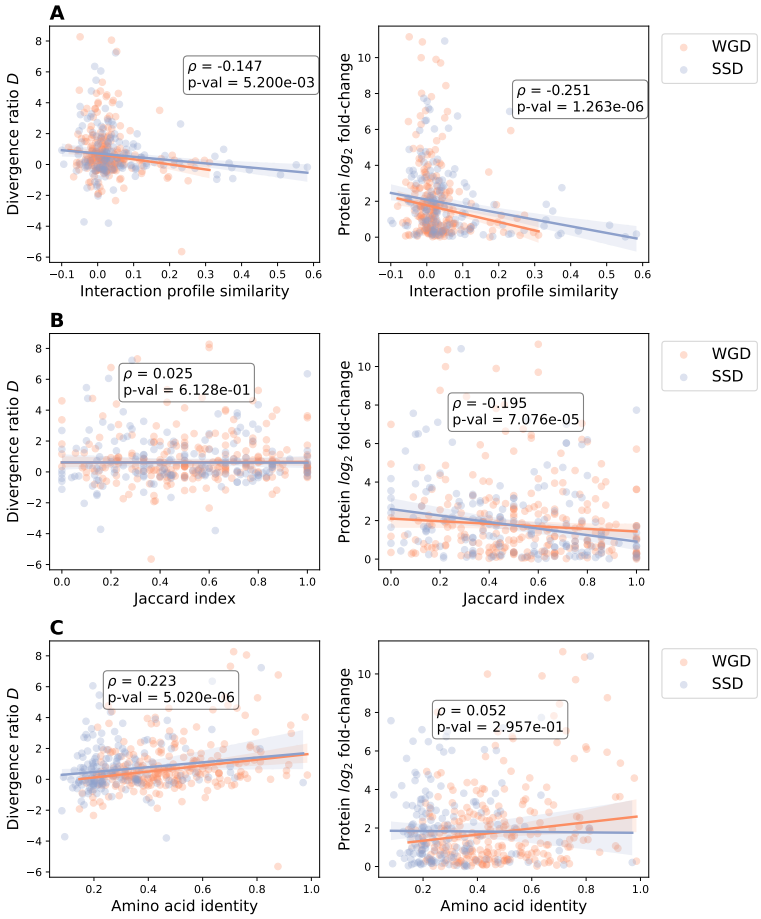

Fig S2. The predominance of transcriptional divergence is only weakly associated with the level of functional divergence among *S. cerevisiae* duplicated genes. (A) Correlation of the divergence ratio (left) or the protein abundance log_2 fold-change (right) with the absolute value of the similarity of genetic interactions profiles within each paralog pair, separately for WGD- and SSD-derived ones. Spearman's ρ and p-value calculated for all duplicate pairs. (B) Correlation of the divergence ratio (left) or the protein abundance log2 fold-change (right) with the overlap (Jaccard index) of GO Slim annotations between paralogs of the same pair, separately for WGD- and SSD-derived ones. Spearman's ρ and p-value calculated for all gene couples. (C) Correlation of the divergence ratio (left) or the protein abundance log_2 fold-change (right) with the amino acid identity between pairs of duplicated genes, separately for WGD- and SSD-derived ones. Spearman's ρ and p-value calculated for all paralog pairs.

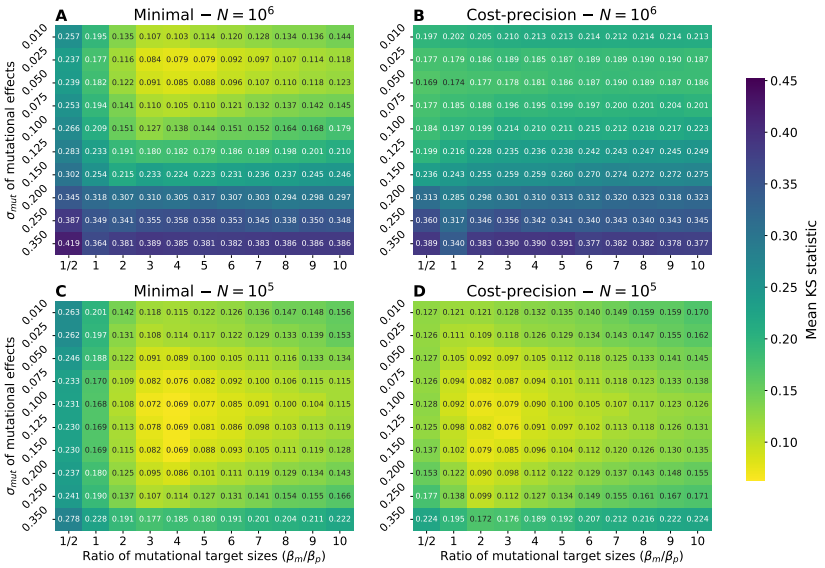

Fig S3. Identification of the best-fitting σ_{mut} across models and selection efficacies. Mean Kolmogorov-Smirnov statistics for the comparisons between empirical and simulated relative divergence distributions are shown. Each value is the grand mean for the four parameters (β_m, β_p, P, and D) across three replicate simulations of 2000 randomly generated paralog pairs, which were stopped according to the protein abundance divergence of yeast WGD-derived paralogs. The lowest values indicate the best fit. (A) Minimal model and high selection efficacy. (B) Cost-precision model and high selection efficacy. (C) Minimal model and reduced efficacy of selection. (D) Cost-precision model and reduced selection efficacy.

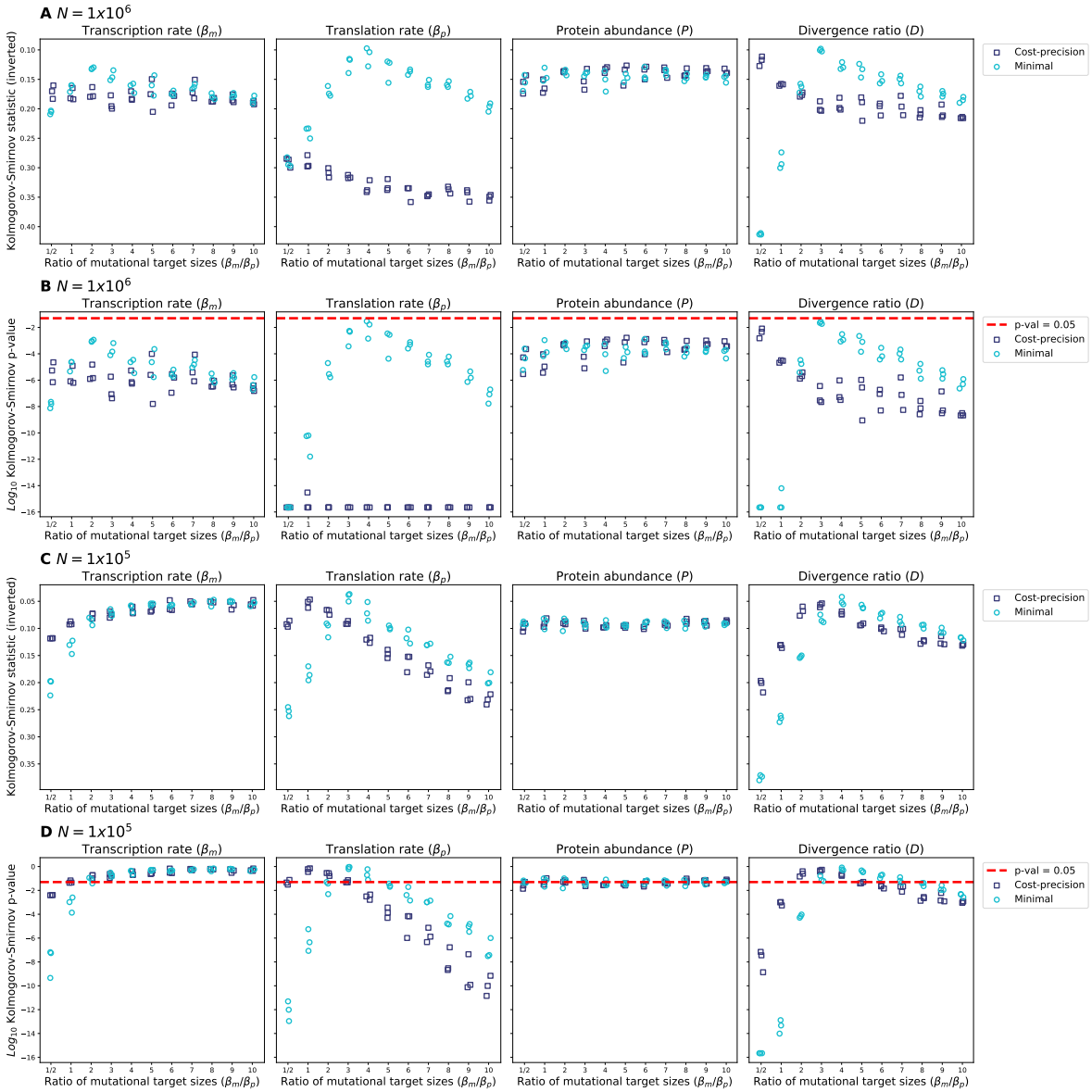

Fig S4. Replication of relative divergence in transcription, translation and protein abundance as well as of the divergence ratio across models and levels of selection efficacy for the corresponding best-fitting σ_{mut} values. Both models can generate realistic distributions for all four levels, but the cost-precision model is unable to replicate the observed translational divergence under high selection efficacy. Simulations shown were stopped according to the protein abundance divergence of yeast WGD-derived paralogs and compared to the empirical divergence patterns of these gene pairs. (A) Kolmogorov-Smirnov statistics for three replicate simulations of 2000 randomly generated paralog pairs across the two models when selection efficacy is high and $\sigma_{mut} = 0.025$. (B) Corresponding log_{10}- transformed p-values of the Kolmogorov-Smirnov test. The red dashed line indicates the threshold above which differences are not significant. (C) Kolmogorov-Smirnov statistics for three replicate simulations of 2000 randomly generated paralog pairs across both models when selection efficacy is lowered and $\sigma_{mut} = 0.100$. (D) Corresponding log_{10}-transformed p-values of the Kolmogorov-Smirnov test. The red dashed line indicates the threshold above which differences are not significant.

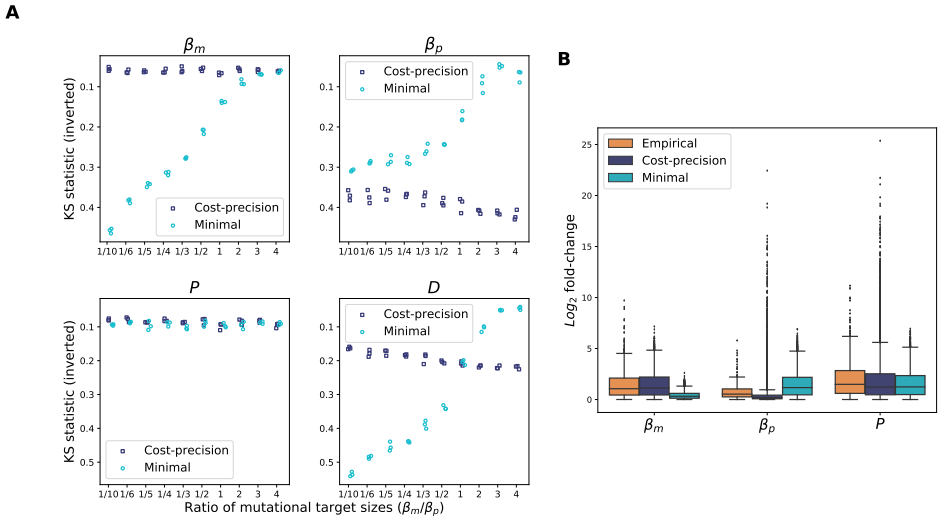

Fig S5. A larger mutational target size for translation does not rescue translational divergence under the cost-precision model when the efficacy of selection is high. (A) Kolmogorov-Smirnov statistics comparing the fit between empirical and simulated relative divergence distributions for transcription rate (β_m), translation rate (β_p) and protein abundance (P) as well as for the divergence ratio D for three replicate simulations of 2000 randomly generated paralog pairs made under the minimal and cost-precision models when selection efficacy is high ($N = 1 \times 10^6$) and $\sigma_{mut} = 0.025$ is postulated. (B) Empirical divergence of WGD-derived yeast paralogs compared with the final simulated divergence in transcription, translation and protein abundance for simulations under the minimal and cost-precision models when the mutational target size of translation is ten times larger than that of transcription ($P_{\beta m}/P_{\beta p} = 1/10$, on panel A) and selection efficacy is high ($N = 1 \times 10^6$; $\sigma_{mut} = 0.025$). All simulations stopped according to the empirical protein abundance relative divergence of *S. cerevisiae* WGD-derived paralogs and compared to the empirical divergence of these gene pairs.

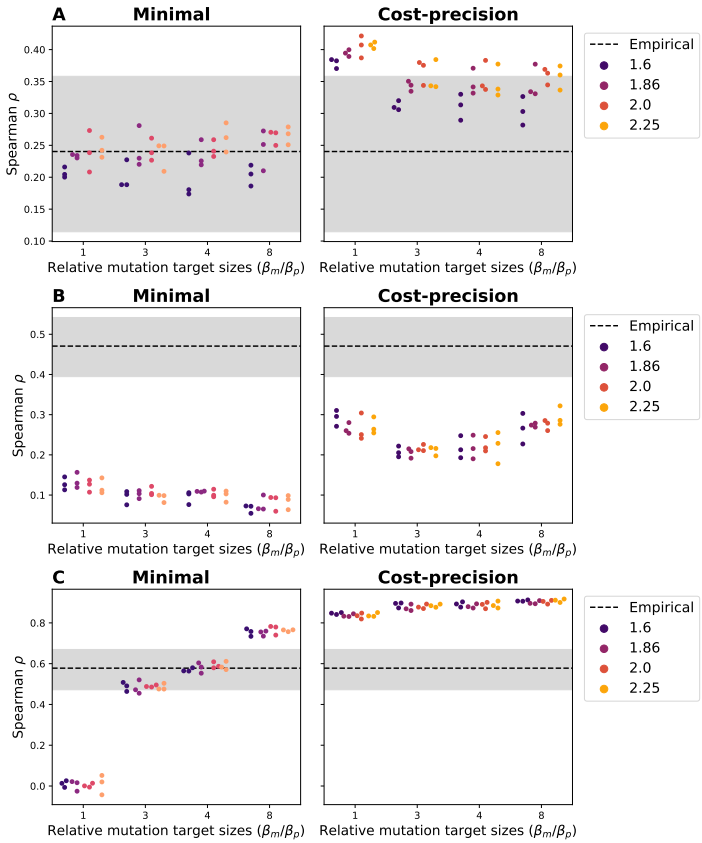

Fig S6. The postulated post-duplication change of optimal cumulative protein abundance has limited influence on the divergence correlations generated from simulations. Simulated divergence correlations for three replicate simulations of 2000 randomly generated paralog pairs under the minimal and cost-precision models for a sample of mutational target size ratios and a range of post-duplication variation Δ_{opt} of the cumulative protein abundance optimum, when the efficacy of selection is assumed to be high ($N = 1 \times 10^6$; $\sigma_{mut} = 0.025$). A Δ_{opt} of 2.0 indicates that the doubling of protein abundance resulting from the duplication event is perfectly optimal. All simulations stopped according to the empirical expression divergence of yeast WGD-derived duplicate pairs. The dashed lines and the shaded areas represent the empirical value of the corresponding divergence correlation and its 95% confidence interval. (A) Correlation between the magnitudes of relative divergence in transcription and translation. (B) Correlation between the signed divergences in transcription and translation. (C) Correlation between the divergence ratio D and the magnitude of relative divergence in protein abundance.

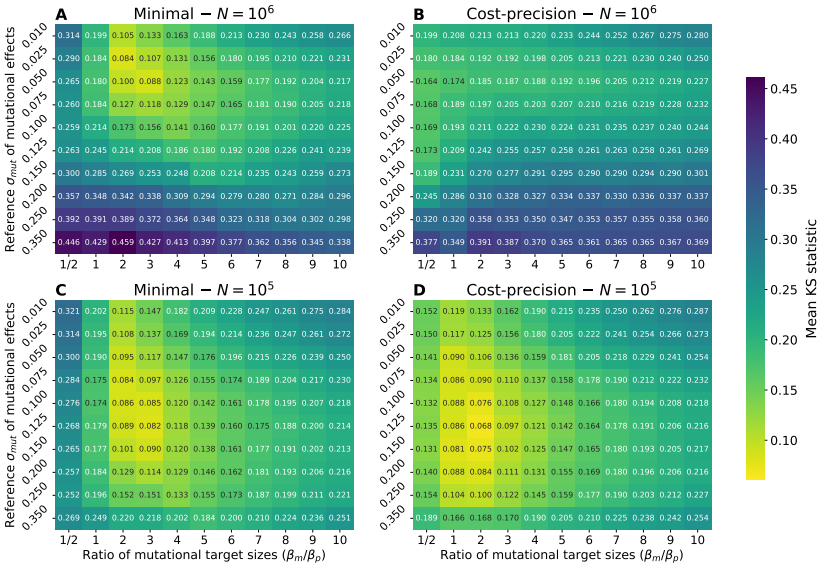

Fig S7. Identification of the best-fitting reference σ_{mut} across models and selection efficacies when bivariate mutational effects are used. Mean Kolmogorov-Smirnov statistics for the comparisons between empirical and simulated relative divergence distributions are shown. Each value is the grand mean for the four parameters (β_m, β_p, P, and D) across three replicate simulations of 2000 randomly generated paralog pairs. The lowest values indicate the best fit. (A) Minimal model and high selection efficacy. (B) Cost-precision model and high selection efficacy. (C) Minimal model and reduced efficacy of selection. (D) Cost-precision model and reduced selection efficacy.

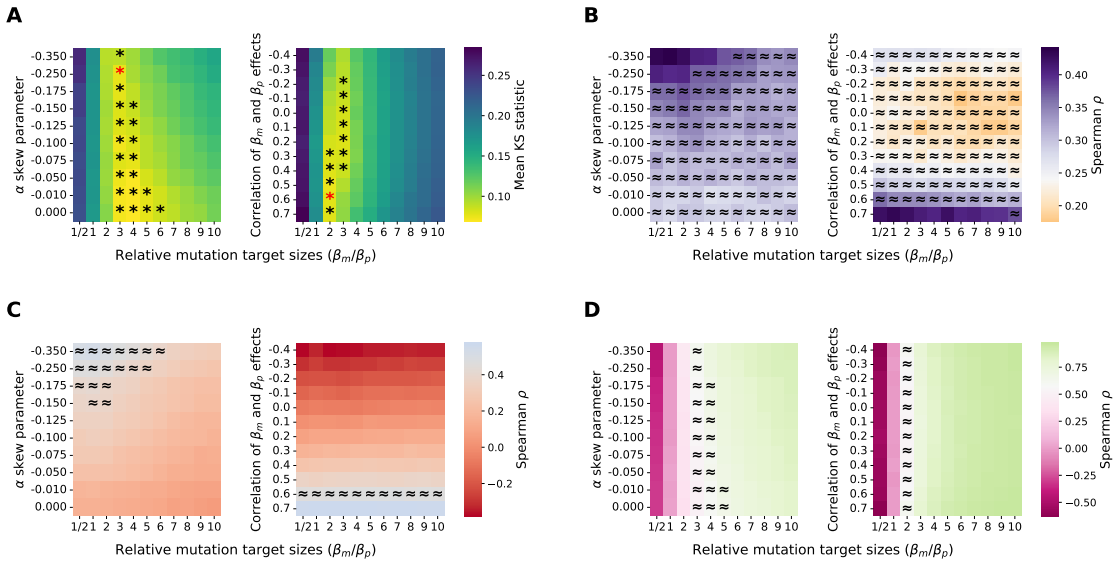

Fig S8. The addition of realistic mutational constraints also allows the minimal model to replicate four features of the empirical divergence of yeast WGD-derived paralogs under conditions of reduced selection efficacy. Results shown for three replicate simulations of 2000 randomly generated paralog pairs performed under the minimal model with the assumption of $N = 1 \times 10^5$ and stopped according to the protein abundance divergence of yeast WGD-derived paralogs. (A) Average replication of the four distributions of relative divergence (transcription, translation, protein abundance and divergence ratio) for a range of mutational effects distribution asymmetry (left) or correlations between the effects of transcriptional and translational mutations (right). Asterisks identify instances where all four magnitudes of relative divergence are realistic for at least one of three replicate simulations (p-val > 0.05, Mood's median test). A red asterisk indicates that this coincides with the recapitulation of all three correlations studied in the subsequent panels. (B, C, D). Average final correlation for three replicate simulations between 1) the magnitudes of transcriptional and translational divergences (B), 2) the signed divergences in transcription and translation (C) as well as 3) the divergence ratio D and the magnitude of protein abundance divergence (D) across the same ranges of mutational effects distribution asymmetry or correlations between the effects of transcriptional and translational mutations. In each case, $\approx$ designates parameter combinations where a correlation coefficient within the 95% confidence interval of the empirical value was obtained for at least one replicate simulation.

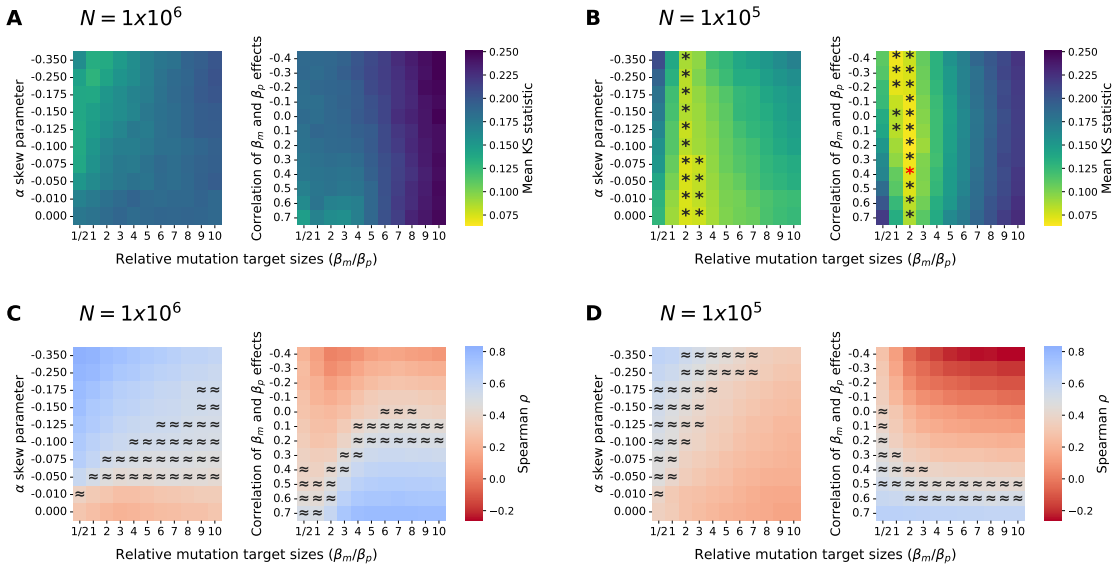

Fig S9. The addition of realistic mutational constraints allows the cost-precision trade-off to simultaneously replicate four features of the empirical divergence patterns of yeast paralogs, but only when the efficacy of selection is reduced ($N = 1 \times 10^5$). Results shown for three replicate simulations of 2000 randomly generated paralog pairs performed under the cost-precision model and stopped according to the protein abundance divergence of yeast WGD-derived paralogs. (A, B) Average replication of the four distributions of relative divergence (transcription, translation, protein abundance and divergence ratio) for a range of mutational effects distribution asymmetry (left) or correlations between the effects of transcriptional and translational mutations (right). Asterisks identify instances where all four magnitudes of relative divergence are realistic for at least one of three replicate simulations (p-val > 0.05, Mood's median test). A red asterisk indicates that this coincides with the recapitulation of all three divergence correlations studied. (C, D) Final correlation between the signed divergences in transcription and translation for the same ranges of mutational effects distribution asymmetry (left) or correlations between the effects of transcriptional and translational mutations (right), averaged across the three replicate simulations. The ≈ symbol identifies parameter combinations where a correlation coefficient within the 95% confidence interval of the empirical value was obtained for at least one replicate simulation.

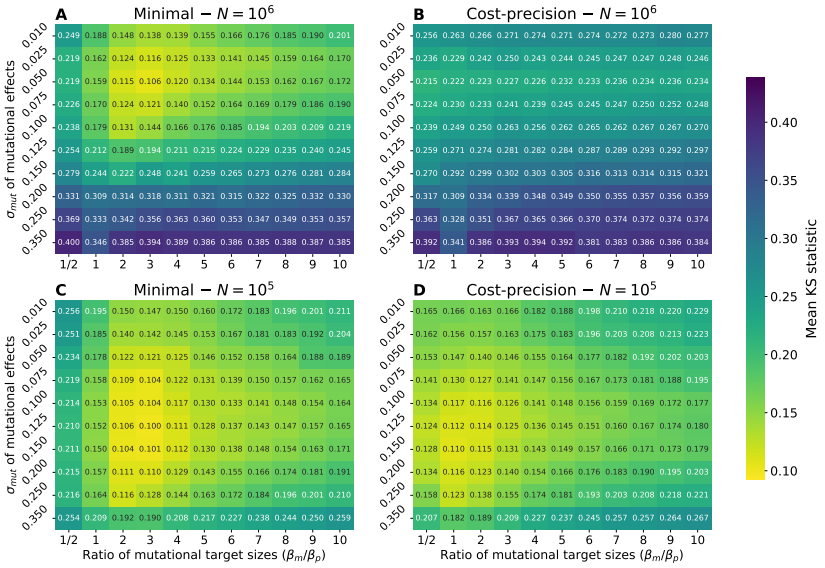

Fig S10. The minimal and cost-precision models offer some replication of the divergence patterns of SSD-derived yeast paralogs, but for distinct parameter combinations and with a lower overall agreement. Grand means of Kolmogorov-Smirnov statistics (across β_m, β_p, P, and D) for three replicate simulations of 2000 randomly generated paralog pairs, each stopped according to the protein abundance divergence of the SSD-derived duplicate pairs of *S. cerevisiae*. (A) Minimal model and high selection efficacy. (B) Cost-precision model and high selection efficacy. (C) Minimal model and reduced efficacy of selection. (D) Cost-precision model and reduced selection efficacy.

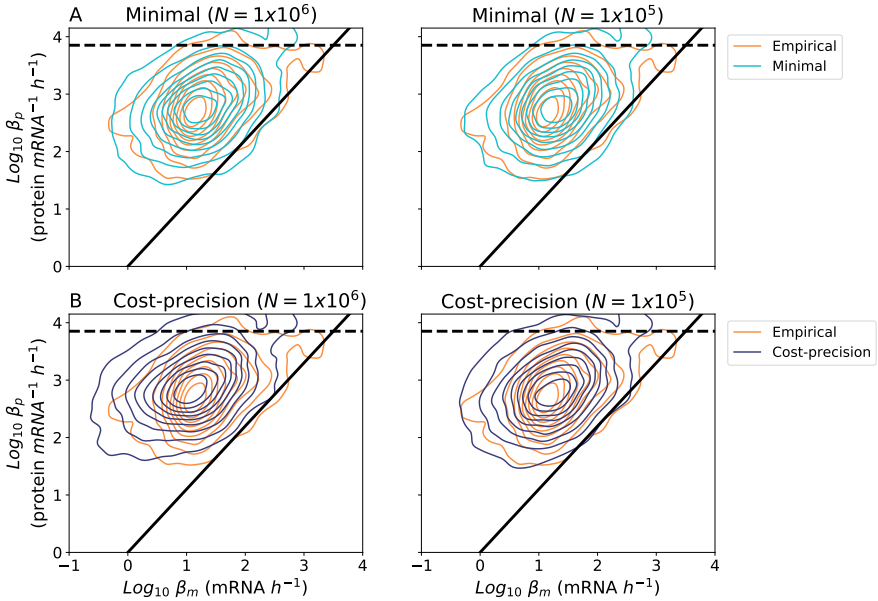

Fig S11. Simulations under both models generate realistic distributions of genes in the expression space. Comparisons of the empirical distribution of yeast genes in transcription (β_m) and translation (β_p) rates [14] with the distributions of simulated paralogs obtained from the combination of three replicate experiments of *in silico* evolution with 2000 randomly generated duplicate pairs. The dashed line represents the estimated maximal translation rate [14], while the diagonal is the boundary of the depleted region defined by Hausser et al. (2019) [14] – below which only 1% of all yeast genes are found. (A) Comparisons with simulations made under the minimal model under high and reduced selection efficacy. Best-fitting simulation (according to the grand means of KS statistics) shown in each case (left, $\sigma_{mut} = 0.025$ and $P_{\beta m}/P_{\beta p} = 4$; right, $\sigma_{mut} = 0.100$ and $P_{\beta m}/P_{\beta p} = 4$). (B) Comparisons with simulations made under the cost-precision model for both levels of selection efficacy. Best-fitting simulation, as identified from the grand means of KS statistics, shown in both cases (left, $\sigma_{mut} = 0.025$ and $P_{\beta m}/P_{\beta p} = 4$; right, $\sigma_{mut} = 0.100$ and $P_{\beta m}/P_{\beta p} = 2$). All simulations stopped according to the protein abundance divergence of *S. cerevisiae* WGD-derived paralogs.