

## Optimizing gene expression by adapting splicing

Idan Frumkin<sup>1,\*,#</sup>, Ido Yofe<sup>1,\*</sup>, Raz Bar-Ziv<sup>1</sup>, Yoav Voichek<sup>1</sup> & Yitzhak Pilpel<sup>1,#</sup>

1- Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel.

\*- Co-first authors

#-correspondence: [frumkin.idan@gmail.com](mailto:frumkin.idan@gmail.com); [pilpel@weizmann.ac.il](mailto:pilpel@weizmann.ac.il);

### Abstract

Can splicing be used by cells to adapt to new environmental challenges? While various adaptation mechanisms for regulating gene expression have been revealed for transcription and translation, the role of splicing and how it evolves to optimize gene-expression patterns has not been thoroughly investigated. To tackle this question, we employed a lab-evolution experimental approach that challenged yeast cells to increase expression levels of a gene that carries an inefficiently-spliced intron. We followed the evolution of multiple lines and found independent routes by which cells adapted. Surprisingly, we did not observe an intron loss event, a mechanism believed to be common in intron evolution. Instead, we identified mutations in *cis* that improved the intron's splicing efficiency and increased the overall expression level of the entire gene. One of these *cis*-acting mutations occurred in an adjacent exon and hampered the functionality of the gene that was not under selection - demonstrating that adaptation of splicing efficiency may sometimes come at the expense of protein activity. Additionally, we observed adaptations in *trans*, which increased the cellular availability of the splicing machinery. These adaptations were achieved either by elevated expression levels of the splicing apparatus or, unexpectedly, by reduced expression levels of other intron-containing genes that are the natural consumers of this process. Ultimately, our work reveals novel molecular means by which the splicing machinery is changed by natural selection to optimize gene-expression patterns of cells.

## Introduction

Changing environments can force cells to change their gene-expression programs, to better accommodate their surroundings. Throughout evolution, cells acquired regulatory mechanisms to tune gene expression, which have been the subject of intensive investigations – focusing mainly on transcription and translation. For example, when cells are challenged to increase protein expression levels, the DNA sequence of genes can change so as to increase transcription<sup>1,2</sup>, support more efficient mRNA translation<sup>3,4</sup>, or result in greater mRNA transcript stability<sup>5,6</sup>. Additionally, the transcription and translation machineries themselves have been shown to adapt to environmental challenges by altering the cellular pools of transcription factors<sup>7</sup> or tRNAs<sup>8,9</sup>.

In evolving expression programs, adaptation often occurs either directly on the genes under pressure (“evolution in *cis*”)<sup>10</sup>, or indirectly, *e.g.* on the expression machineries, mostly transcription and translation (“evolution in *trans*”)<sup>11,12</sup>. These two routes of evolution are profoundly different, as the first (*cis*) provides a localized solution that in principle can affect only a certain gene, while the later (*trans*) could be the method of choice if a coordinated change in many genes is needed.

Surprisingly, although the process of splicing is central to the maturation and regulation of mRNAs in eukaryotes<sup>13–17</sup>, its role in adapting to novel demands on gene expression has not been thoroughly investigated. During mRNA splicing, precursor mRNAs are processed to remove introns while fusing exons together to create the mature transcript. This process provides an evolutionary means to diversify the proteome towards phenotypic novelty, as the choice of intron to be excluded, as well as the exons which are found in the mature transcript, can both be regulated based on the cell’s needs<sup>15,18,19</sup>. One aspect of splicing evolution that has been extensively studied is gain and loss of intronic DNA, for which several molecular models have been proposed, mainly Reverse-Transcription and recombination-mediated intron loss, intron transposition and also exonization and intronization via mutations<sup>20–23</sup>. While intron loss and gain have been demonstrated experimentally<sup>24,25</sup>, other forms of splicing evolution, such as alterations in splicing efficiency under changing conditions, have not.

Here, we set out to reveal whether introns or the splicing apparatus can evolve so as to alter the expression levels of genes in a timely and adaptive manner, and ask whether and how splicing evolves in *cis* and in *trans* to regulate gene expression. To this end, we generated a reporter construct in yeast cells that could simultaneously be read out and be selected for splicing efficiency. Namely, we introduced an inefficiently-spliced intron to a reporter gene that was fused to an antibiotic resistance gene. Using this approach, we could carry out a lab-evolution experimental setup to study the adaptation of splicing in the presence of the corresponding antibiotics.

Our results demonstrate two alternative adaptive routes for this evolutionary challenge. First, *cis*-acting solutions, in the form of adaptive mutations, occurred in the intron itself, but also surprisingly in an up-stream exon. These mutations resulted in increased splicing efficiency and higher expression levels of the antibiotic resistance gene. Remarkably, in some evolutionary lines there were no *cis* mutations, but rather *trans*-acting adaptations that have increased cellular availability of the splicing machinery - either by increased expression levels of the machinery's genes themselves, or surprisingly, by decreasing the expression levels of other intron-containing genes. Thus, this work unravels different layers at which the splicing machinery can be adapted to alter gene expression.

## Results

### Low splicing efficiency leads to stressed cells under restrictive conditions

We hypothesized that tuning splicing of genes could serve as a means to optimize their expression levels. To test this hypothesis, we used the yeast *Saccharomyces cerevisiae* in which ~30% of the transcriptome must be spliced, at a range of splicing efficiencies<sup>17,26</sup>, to form mature mRNAs<sup>27</sup>. We built a synthetic gene construct that consists of two fused domains: A fluorescent reporter (YFP), which includes two alternative natural introns - with either high or low splicing efficiency - near the YFP's fluorescence site<sup>26</sup>, fused to an antibiotics resistance gene (Kanamycin resistance gene). Specifically, we created three strains: (i) WT YFP strain without an intron; (ii) "Splicing<sup>High</sup>" in which the YFP harbors the natural intron of *OSH7* and was previously reported to have high splicing efficiency within this YFP context<sup>26</sup>; and (iii) "Splicing<sup>Low</sup>" in which the natural intron of *RPS26B*, with a low splicing efficiency<sup>26</sup>, was inserted in the same location (Figure 1A).

We first hypothesized that cellular growth of each strain in the presence of the antibiotics, geneticin (G418), will associate with the YFP-Kan expression levels. We followed the growth of the three strains in the presence of the antibiotics and found that the WT strain had the highest fitness, Splicing<sup>High</sup> grew slower, and Splicing<sup>Low</sup> demonstrated a severe growth defect compared to the two other strains (Figure 1B+C). We then measured fluorescence intensity of the YFP-Kan reporter in the presence of the drug. In line with the growth measurements, we observed that WT cells demonstrated the highest fluorescence levels, followed by Splicing<sup>High</sup>, and with Splicing<sup>Low</sup> cells showing the lowest YFP-Kan levels (Figure 1D). These results demonstrate that the inefficiently-spliced intron in Splicing<sup>Low</sup> reduces cellular levels of YFP-Kan and hence lead to a reduced fitness.

Since YFP-Kan expression level in Splicing<sup>Low</sup> were significantly lower compared to the other strains, we hypothesized that Splicing<sup>Low</sup> cells did not reach the needed concentration to sufficiently neutralize the antibiotics, and hence resulted in stressed

cells. To test this hypothesis, we performed mRNA sequencing of exponentially growing WT and Splicing<sup>Low</sup> cells in an antibiotics containing medium, and analyzed their transcriptome profiles. Indeed, we observed that ribosomal genes were down-regulated in Splicing<sup>Low</sup> compared to the control strain – a clear signature of stressed cells<sup>28</sup> (Figure 1E). Notably, the reduction in ribosomal expression levels (~8%) we observed here due to growth rate differences between WT and Splicing<sup>Low</sup> cells is accurately predicted by a recent study, which calculated the linear correlation between growth rate and ribosomal expression levels in yeast cells<sup>29</sup>. In parallel, stress-related genes<sup>30</sup> were up-regulated in the Splicing<sup>Low</sup> compared to the control strain (Figure 1E). We thus concluded that the general stress response was activated in Splicing<sup>Low</sup> cells.

### **Rapid evolutionary adaptation increases expression level of the resistance gene**

Our experimental system mimics an evolutionary scenario in which there is an immediate and continuous selection pressure to up-regulate the expression level of a gene of interest. How would the system evolve to better resist the antibiotics? Possible means to adapt include mutations in the gene's promoter to increase transcription, mutations that increase translation initiation, or mutations inside the gene itself that increase the functional efficiency of the protein. Additionally, the splicing machinery may also take part in adaptation of gene expression levels. To find which evolutionary track would be used by cells, we evolved the three strains by daily serial dilution on a medium supplemented with G418 for ~560 generations, in four independent cultures for each strain (Figure 2A). Interestingly, only the cultures of Splicing<sup>Low</sup> cells demonstrated a significant improvement in fitness at the end of the experiment (Figure 2B+C). This observation implies that only Splicing<sup>Low</sup> experienced a sufficiently strong selective pressure to adapt to the presence of the antibiotics in the medium, in contrast to the WT and Splicing<sup>High</sup> strains which originally had much higher levels of the resistance gene.

Consistent with the fitness measurements, YFP measurements of the evolved cultures showed that expression levels of the resistance-YFP fusion gene increased in all four evolved cultures of Splicing<sup>Low</sup> compared to the ancestral strain (Figure 2D). Conversely,

the increase in YFP-Kan expression levels in the evolved WT populations was smaller, and only one culture of the evolved Splicing<sup>High</sup> cells demonstrated strong elevation of the YFP-Kan levels (Figure 2D). These results further support that Splicing<sup>Low</sup> cells experienced the strongest selective pressure to adapt rapidly to the presence of the antibiotics in our experimental setup, and that they achieved this goal by increasing the levels of the resistance gene. We next moved to reveal the molecular mechanisms underlying this evolutionary process.

### **Adaptation in *cis* and *trans* leads to increased splicing efficiency**

We hypothesized that improving the low splicing efficiency of the intron in Splicing<sup>Low</sup> could be exploited by natural selection as an adaptation mechanism towards increasing the resistance gene levels. We therefore sequenced the YFP-Kan locus in 16 randomly chosen colonies from two evolved populations (termed here population A and population B) of Splicing<sup>Low</sup>. Interestingly, we found that the colonies were split into two types – either with or without a mutation in the YFP-Kan locus. In population A, we found that the same mutation occurred in four out of eight colonies, changing adenine to cytosine inside the intron, 97 nucleotides up-stream to its 3' end (Figure 3A). In population B, we identified an exonic non-synonymous mutation that changed a valine at position 61 of the YFP protein into alanine (a thymine to cytosine 14 nucleotides up-stream of the intron) in three out of eight colonies. In the five other colonies from this population there were no mutations in the YFP-Kan locus.

Notably, none of the colonies demonstrated a mutation in the construct's promoter, terminator or in the sequence of the Kan resistance gene itself. These results propose that different mutations in the intron, or its vicinity, were adaptive and might affect splicing efficiency of the intron. Surprisingly, the observed mutations did not occur in the 5' donor, 3' acceptor, nor in the intron branch point – suggesting that other position of the intron can also be selected in evolution increase fitness by affecting splicing. While the intron- and exon-mutated colonies represent an evolutionary adaptation in *cis*, the

colonies that showed no mutation in the entire gene construct potentially found adaptive solutions in *trans* that may have occurred elsewhere in the genome.

We randomly chose six colonies: four colonies with a *cis* mutation and two colonies that showed no mutations in *cis*, for which we reasoned that such colonies may have adapted in *trans*. We termed these colonies according to the evolution lines from which they were derived: A-*cis*1, A-*cis*2, B-*cis*1, B-*cis*2, A-*trans* and B-*trans*. We followed the growth of these evolved colonies in the presence of G418 and found, as expected, that all grew faster than the Splicing<sup>Low</sup> ancestor (Figure 3B). We then performed RNA-seq and transcriptome analysis of all colonies, which revealed relaxation of the stress response that was featured in the ancestor. Namely, the general stress response genes were reduced and ribosomal proteins were up-regulated in five evolved colonies (Figure 3C and Supplementary Figure 1). These observations suggest that the cells indeed adapted to the presence of the antibiotics in the environment and that the stress experienced by them was partially alleviated.

We next hypothesized that cellular fitness might correlate with mRNA levels of the YFP-Kan construct because increased transcript levels should result in higher concentrations of the Kan protein. Indeed, maximal growth rates of the control and Splicing<sup>Low</sup> ancestors and for the six evolved colonies correlate with mRNA levels of the YFP-Kan construct, as deduced from the RNA-seq (Figure 3D) – supporting our conclusion that adaptation was based on increasing expression levels of the YFP-Kan gene. Since the observed *cis* mutations occurred at the vicinity of the intron, we hypothesized that they increased splicing efficiency of the YFP-Kan transcript. To test this possibility, we performed, for both *cis*- and *trans*-evolved colonies, a splicing efficiency assay with qPCR - targeting the un-spliced and spliced transcript versions. Interestingly, the ratio of spliced to un-spliced transcripts was higher in all evolved colonies compared to the Splicing<sup>Low</sup> ancestor, suggesting that at least some of the mRNA level increase we observed in the evolved colonies results from increased splicing efficiency (Figure 3E).

To prove that adaptation of the colonies actually led to higher protein levels of the resistance gene, we measured fluorescence intensity using flow cytometry. We found that the two *cis*-colonies from population A (A-*cis*1 & A-*cis*2) and the two *trans*-colonies (A-*trans* & B-*trans*) showed higher YFP-Kan levels compared to the ancestor. However, the two *cis*-colonies from population B (B-*cis*1 & B-*cis*2) demonstrated decreased fluorescence intensity values (Figure 3F). These observations suggest that the non-synonymous, exon mutation reduced the fluorescence-per-protein value of the YFP-Kan construct in these colonies. Indeed, this position corresponds to a position that was recently reported to reduce fluorescence when mutated in the highly similar GFP<sup>31</sup>. Because YFP functionality was not selected for or against in our setup, it was free to mutate as long as it helps achieve a higher expression level of the entire construct by increasing the intron's splicing efficiency. It thus seems that modular domain-architecture of a protein may increase its evolvability under relevant conditions as it allows the optimization of each domain in isolation from the other.

It is possible that additional beneficial mutations exist in the genome of the *cis*-evolved colonies, which account for the phenotypes we observed. To directly assess the effects of the *cis* mutations, we generated two rescue strains, termed rescue-A and rescue-B, in which these *cis*-acting mutations were introduced individually to the ancestral Splicing<sup>Low</sup> background. Notably, the two rescue strains grew better than Splicing<sup>Low</sup> cells in the presence of the antibiotics (Figure 4A), though not as good as the wild-type, and the stress experienced by the Splicing<sup>Low</sup> cells was relieved upon insertion of each individual *cis* mutation (Figure 4B). Finally, we measured splicing efficiencies and fluorescence intensity levels for both rescue strains, and found that they resembled the results of the evolved single colonies (Figure 4C-D, in comparison to Figure 3E-F). These observations strengthen our conclusions that the *cis*-acting mutations are sufficient to elevate YFP-Kan levels through an increased splicing efficiency, yet the non-synonymous mutation of population D also hampers the function of the YFP domain and reduces its fluorescence-per-protein ratio.



Our results thus far provide direct evidence that intron splicing takes part in the adaptation and optimization of gene expression patterns to environmental needs. Although intron sequences are much less conserved compared to exons, and are believed to be less functional, we demonstrate that their sequence can be used by natural selection as a molecular mechanism to regulate splicing efficiency and adjust gene expression patterns.

### **Increasing cellular availability of the splicing machinery can be adaptive**

We finally aimed to decipher the mechanism behind the increased YFP-Kan levels in the *trans*-evolved colonies that showed no mutations in *cis*, i.e. within the reporter gene or in its vicinity. We reasoned that elevating availability of the splicing machinery as a global resource could be a means to increase splicing efficiency of the YFP-Kan transcript, and thus could be used as an adaptive mechanism to the antibiotics challenge. Increased splicing-availability could be achieved by increasing the expression of the splicing machinery genes. In addition, as with other cellular machineries whose functioning depends on supply-to-demand economy<sup>4,8,32,33</sup> reducing expression levels of the intron-containing genes, namely the “demand”, could increase the availability of the machinery towards the intron under selection here.

To test if any of these evolutionary routes were indeed taken by the evolved cells, we calculated the expression level ratio of genes between the evolved colonies and their ancestor. In colony A-*trans*, we observed that while the average expression-ratio of the splicing machinery genes (the “supply”) increased, that of the non-ribosomal intron-containing genes (the “demand”) decreased (Figure 5A). This observation suggests that indeed the cellular availability of the splicing machinery was elevated in this evolved colony, which might have allowed for the observed increased splicing efficiency of the YFP-Kan gene. Next, we hypothesized that the *cis*-evolving colonies may have also adapted in *trans* and used this adaptation mechanism as well. Indeed, in all other evolved colonies we observed a similar trend, in which the overall supply-to-demand measurement of the splicing machinery was increased (Figure 5B). Notably, in some

colonies this phenotype was achieved by only increasing expression levels of splicing genes or by only reducing levels of the intron-containing genes (Supplementary Figure 2). Importantly, the two rescue strains, which did not evolve and only harbor our artificially introduced *cis*-acting mutation, did not show any change in splicing availability (Figure 5B), strongly supporting our conclusion that this phenotype was achieved by further adaptation of the cells during our lab-evolution experiment. Thus, we concluded that both *cis* and *trans* adaptation routes can co-occur in the same genome towards optimization of its gene expression patterns.

## Discussion

Here we study the role of the splicing machinery in optimization of gene expression programs by placing selective pressure on cells to improve the splicing efficiency of a specific gene. Our results provide molecular evidence for the relevance of splicing as another instrument in the cellular toolbox towards adjusting its gene expression patterns. To the best of our knowledge, we demonstrate the first experimental evidence of splicing efficiency adaptation, confirming that this adaptation can occur in *cis* and *trans* similarly to adaptations of other means of gene regulation.

Two potential solutions to the burden we imposed on our ancestor lines were, surprisingly, not realized during our lab-evolution. First, considering previous studies of splicing evolution, one could have expected the intron to be lost by a genomic deletion or reverse transcription<sup>34,35</sup>. Such a solution could have been an ideal evolutionary adaptation to alleviate the burden, as we show that the intron-less strain has the highest fitness. The fact that we did not observe an intron-loss event suggests that nucleotide substitutions were more accessible solutions in this case, in agreement with previous evidence in yeast that nucleotide mutations are much more prevalent than deletion events, at a ratio of 33:1<sup>36</sup>. Another surprise was that the mutations we observed did not occur within any of the intron's three functional sites: the 5' donor, 3' acceptor, or the branch point of splicing. Surprisingly, one mutation was actually detected, and was verified here to affect splicing, in a region of the intron not known to exert a major effect on splicing, and another splicing-improving mutation happened in the up-stream exon - suggesting that various positions in the intron and its proximity may facilitate splicing rate and take part in the evolution of this process.

Notably, the fluorescence intensity per protein molecule of the YFP domain was decreased due to the non-synonymous mutation in the YFP first exon, suggesting that under certain evolutionary constraints selection may hamper superfluous functions of certain protein domains so as to increase availability of the entire protein. Why then, would the mutations we observed increase splicing efficiency? Past evidence showed that

mRNA secondary structures at the intron's edges influence splicing efficiency<sup>26,37-39</sup>. It is possible then that the mutations we observed in this study somehow open the structure of the intron under selection and make it more accessible for the splicing machinery.

Adaptive changes also occurred in *trans* to the YFP-Kan locus and increased availability of the splicing machinery. Recently, the competition of pre-mRNAs for the splicing machinery was shown to affect cellular function, as splicing efficiency of multiple introns was influenced by changes in the composition of the transcript pool<sup>40</sup>. While this mechanism was elegantly suggested to maintain the separation between meiotic and vegetative gene expression states, it is also possible that it can be utilized as an adaptive route available for cells to optimize expression levels of genes. More broadly, it has been shown<sup>41</sup> that yeast species that have a high content of intron-containing genes have adapted the codon usage of their splicing machinery genes more vigorously to their cellular tRNA pools compared to other species that have a lower number of introns in the genome. This evolutionary trend indeed shows that the splicing machinery adaptively responds to meet its own evolutionary demand.

Our findings demonstrate how availability of the splicing apparatus may have been beneficially increased both by elevating the expression level of the machinery and/or by reducing other intron-containing genes that compete with the antibiotic resistance unspliced RNA for the spliceosome. Thus, increase in supply-to-demand ratio, analogous to the case in translation systems<sup>8,42</sup>, appears to have evolved in this case.

Interestingly, we found that different adaptive means co-occurred in the evolved populations – independently in different cells or even simultaneously in the same genome. In particular, we saw that evolutionary lines that adapted in *cis* appear to also have had adaptations that are not encoded in the evolving gene, hence pointing to changes that must have occurred in *trans*. Further investigations will reveal which of these solutions, *cis* or *trans*, proves to be more evolutionarily stable - to fully reveal the dynamics of splicing adaptation when cells optimize their gene expression.

## Materials and Methods

### Yeast strains and plasmids

All *S. cerevisiae* strains in this study have the following genetic background: his3 $\Delta$ 1::TEF2-mCherry::URA3::RPS28Ap-YiFP-KAN::NAT; can $\Delta$ 1::STE2pr-Sp\_his5; lyp $\Delta$ 1::STE3pr-LEU2; leu2 $\Delta$ 0; ura3 $\Delta$ 0;

Strains of Y-intron-FP were taken from Yofe *et al*<sup>26</sup> and were introduced with a Kan resistance gene fused 3' terminally to the YFP. To reconstitute the mutations discovered after lab evolution (rescue strains), we amplified cassettes of Y-i<sub>mut</sub>-FP-KAN and transformed these into the ancestor WT strain, selecting with KAN. Notably, all strains also carry an mCherry-fluorescent protein driven by an independent *TEF2* promoter that was used to normalize cell-to-cell variability for the YFP-Kan expression levels.

### Media

Cultures were grown at 30°C in rich medium (1% bacto-yeast extract, 2% bacto-peptone and 2% dextrose [YPD]). Throughout all experiments, G418 was supplemented to the medium at a concentration of 3mg/ml, which is 10fold higher than the standard.

### Evolution experiments

Lab-evolution experiments were carried out by daily serial dilution for 80 days. Cells were grown on 1.2ml of YPD+G418 at 30°C until reaching stationary phase and then diluted by a factor of 1:120 into fresh media (~7 generations per dilution, total of ~560 generations).

### Liquid growth measurements

The cultures were grown at YPD+G418, and optical density (600nm) measurements were taken at 30min intervals. Growth comparisons were performed using 96-well plates, and the growth curve for each strain was obtained by averaging at least 15 wells.

### **FACS measurements of YFP-Kan levels**

Cells were grown in YPD+G418 at 30°C until they reached the logarithmic growth phase at an optical density of ~0.4. Then, YFP and mCherry levels were measured for ~50,000 cells for each culture with flow cytometry. Gating was performed according to side and forward scatters, and YFP levels were normalized with the mCherry signal for each cell individually.

### **Quantitative PCR measurements of splicing efficiency**

Cultures were grown in YPD+G418 at 30°C until cells reached the logarithmic growth phase at an optical density of ~0.4. Then, RNA was extracted using MasterPure kit (Epicentre), and were reverse-transcribed to cDNA using random primers. 2µl of cDNA were added to each reaction as template for qPCR using light cycler 480 SYBR I master kit and the LightCycler 480 system (Roche Applied Science), according to the manufacturer's instructions. For each strain, two qPCRs were performed with three biological repetitions and three technical repeats. A first qPCR was performed targeting the transcript spliced-version with a forward primer complementing the exon-exon junction and a downstream reverse primer. A second reaction targeted the un-spliced version of the transcript with a forward primer complementing the intron and the same reverse primer of the first reaction.

$F_{\text{exon-exon}} = 5' \text{-CACTACTTTAGGTTATGGTTT-3}'$

$F_{\text{intron}} = 5' \text{-CTTCAATTTACTGAATTTGTATG-3}'$

$R_{\text{both}} = 5' \text{-GTCTTGTAGTTACCGTCA-3}'$

Splicing efficiency is reported as the average  $C_p$  of the spliced transcript minus the average  $C_p$  of the un-splice version.

### **mRNA deep sequencing**

Cultures were grown in YPD+G418 at 30°C until cells reached the logarithmic growth phase at an optical density of ~0.4. Cells were then harvested by centrifugation and flash-frozen in liquid nitrogen. RNA was extracted using a modified protocol of nucleospin® 96

RNA kit (Machery-Nagel). Specifically, cell lysis was done in a 96 deep-well plate by adding for each well 450 $\mu$ l of lysis buffer containing 1M sorbitol, 100mM EDTA and 0.45 $\mu$ l lyticase (10IU/ $\mu$ l). The plate was incubated at 30°C for 30min to break cell wall, centrifuged for 10min at 3000rpm, followed by the removal of the supernatant. Then, extraction continued as in the protocol of nucleospin® 96 RNA kit, only using  $\beta$ -mercaptoethanol instead of DTT. Poly(A)-selected RNA extracts of size ~200bps were reverse-transcribed to cDNA using poly(T) primers that were barcoded with a unique molecular identifier (UMI). cDNA was then amplified and sequenced with an Illumina HiSeq 2500.

### **Analysis of mRNA deep sequencing**

Processing of RNA-seq data was performed as described in Voichek *et al*<sup>43</sup>. Shortly, reads were aligned using Bowtie<sup>44</sup> (parameters: --best -a -m 2 -strata -5 10) to the genome of *S. Cerevisiae* (R64 from SGD) with an additional chromosome containing the sequence of the YFP-Kan construct. For each sequence, we normalized for PCR bias using UMIs as describe in Kivioja *et al*<sup>45</sup>. Next, reads for each gene end (400bp upstream to 200bp downstream of the ORF's 3' end) were summed-up to estimate the gene's expression level. Genes with coverage lower than 10 reads were excluded. To normalize for differences in coverage among samples, we divided each gene expression by the total read count of each sample and then multiplied by 10<sup>6</sup>. Then, expression ratio was calculated between an evolved/rescue colony to the ancestor and a log2 operation was performed on that ratio. These values were used to compare expression levels of gene groups (ribosomal genes, general stress response genes, splicing machinery genes, intron-containing genes) and of the YFP-Kan mRNA levels as described in the manuscript. When calculating the expression levels of splicing machinery and intron-containing gene groups, the ribosomal and general stress response genes were excluded from the analysis in order to avoid bias from cellular regulation due to changes in physiology and growth rate of the cells.

## **Acknowledgments**

We thank Martin Mikl for help with the qPCR experiments. We also greatly appreciate comments to the text from Tsilil Ast, Orna Dahan, Dvir Schirman, and Avihu Yona, and the entire Pilpel lab for discussions of the project. IF thanks the Azrieli Foundation for an Azrieli PhD Fellowship. This study was supported by the Minerva Foundation who funded the Minerva Center for Live Emulation of Evolution in the Lab.

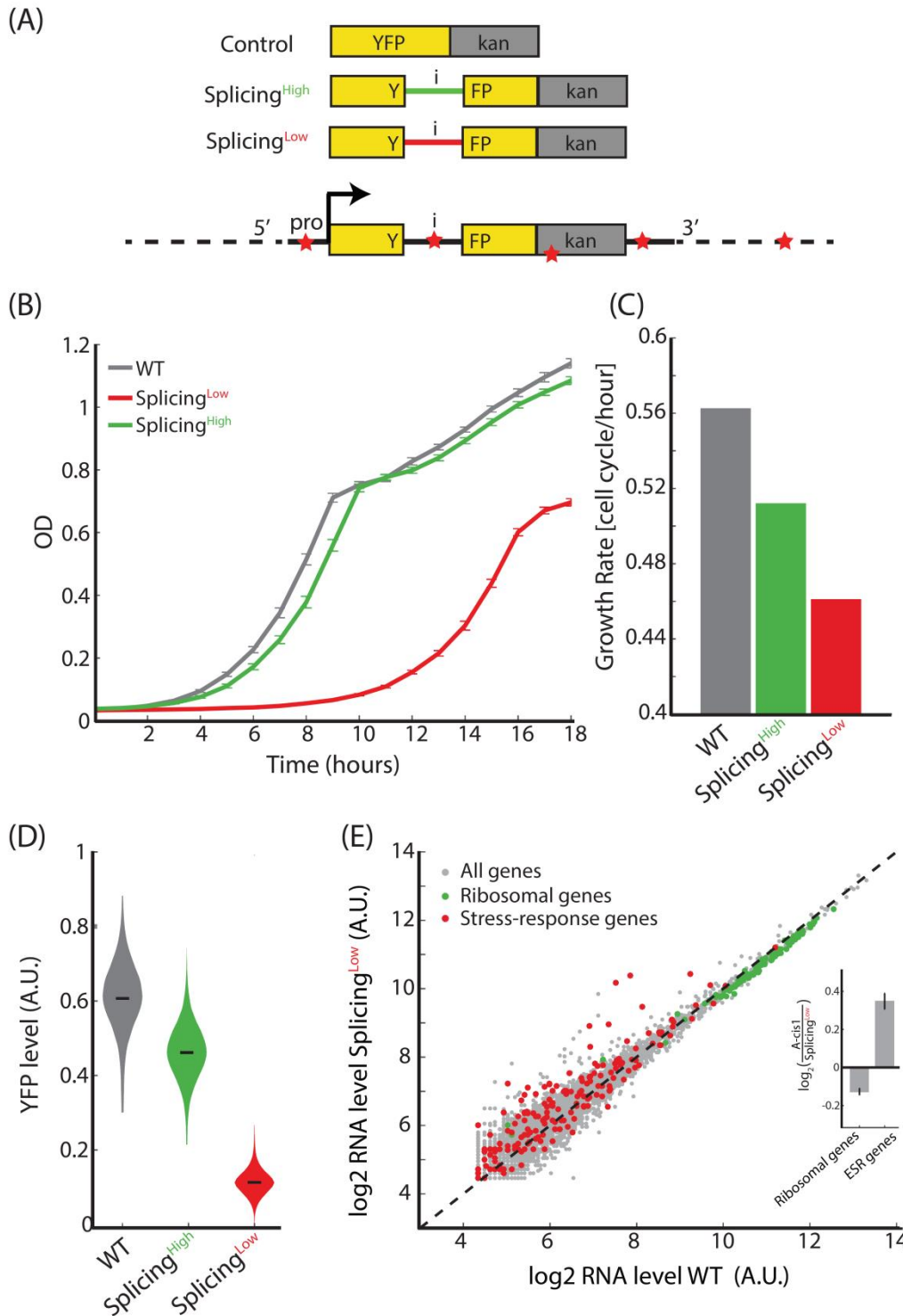


## Bibliography

1. Weingarten-Gabbay, S. & Segal, E. The grammar of transcriptional regulation. *Hum. Genet.* (2014). doi:10.1007/s00439-013-1413-1
2. Amorós-Moya, D., Bedhomme, S., Hermann, M. & Bravo, I. G. Evolution in regulatory regions rapidly compensates the cost of nonoptimal codon usage. *Mol. Biol. Evol.* **27**, 2141–51 (2010).
3. Quax, T. E. F., Claassens, N. J., Söll, D. & van der Oost, J. Codon Bias as a Means to Fine-Tune Gene Expression. *Mol. Cell* **59**, 149–161 (2015).
4. Gingold, H. & Pilpel, Y. Determinants of translation efficiency and accuracy. *Mol. Syst. Biol.* **7**, 481 (2011).
5. Presnyak, V. *et al.* Codon optimality is a major determinant of mRNA stability. *Cell* **160**, 1111–24 (2015).
6. Bazzini, A. A. *et al.* Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *EMBO J.* **35**, 2087–2103 (2016).
7. Babu, M. & Teichmann, S. A. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res.* **31**, 1234–44 (2003).
8. Yona, A. H. *et al.* tRNA genes rapidly change in evolution to meet novel translational demands. *Elife* **2**, e01339 (2013).
9. Kirchner, S. & Ignatova, Z. Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nat. Rev. Genet.* (2014). doi:10.1038/nrg3861
10. Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–16 (2007).
11. Lemos, B., Araripe, L. O., Fontanillas, P. & Hartl, D. L. Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14471–6 (2008).
12. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85–8 (2004).
13. Matera, a. G. & Wang, Z. A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15**, 108–121 (2014).
14. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **338**, 1593–9 (2012).
15. Barbosa-Morais, N. L. *et al.* The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**, 1587–93 (2012).
16. Petibon, C., Parenteau, J., Catala, M. & Elela, S. A. Introns regulate the production of ribosomal proteins by modulating splicing of duplicated ribosomal protein genes. *Nucleic Acids Res.* **44**, gkw140 (2016).

17. Parenteau, J. *et al.* Introns within ribosomal protein genes regulate the production and function of yeast ribosomes. *Cell* **147**, 320–31 (2011).
18. Reyes, a. *et al.* Drift and conservation of differential exon usage across tissues in primate species. *Proc. Natl. Acad. Sci.* 2–7 (2013). doi:10.1073/pnas.1307202110
19. Bush, S. J., Chen, L., Tovar-Corona, J. M. & Urrutia, A. O. Alternative splicing and the evolution of phenotypic novelty. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **372**, (2017).
20. Irimia, M. & Roy, S. W. Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb. Perspect. Biol.* **6**, (2014).
21. Roy, S. W. & Irimia, M. Mystery of intron gain: new data and new models. *Trends Genet.* **25**, 67–73 (2009).
22. Hooks, K. B., Delneri, D. & Griffiths-Jones, S. Intron evolution in Saccharomycetaceae. *Genome Biol. Evol.* (2014). doi:10.1093/gbe/evu196
23. Shabalina, S. a *et al.* Distinct patterns of expression and evolution of intronless and intron-containing mammalian genes. *Mol. Biol. Evol.* **27**, 1745–9 (2010).
24. Lee, S. & Stevens, S. W. Spliceosomal intronogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 201605113 (2016). doi:10.1073/pnas.1605113113
25. Derr, L. K., Strathern, J. N. & Garfinkel, D. J. RNA-mediated recombination in *S. cerevisiae*. *Cell* **67**, 355–64 (1991).
26. Yofe, I. *et al.* Accurate, Model-Based Tuning of Synthetic Gene Expression Using Introns in *S. cerevisiae*. *PLoS Genet.* **10**, e1004407 (2014).
27. Ares, M., Grate, L. & Pauling, M. H. A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* **5**, 1138–9 (1999).
28. de Nadal, E., Ammerer, G. & Posas, F. Controlling gene expression in response to stress. *Nat. Rev. Genet.* **12**, 833–45 (2011).
29. Metzl-Raz, E. *et al.* Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *Elife* **6**, (2017).
30. Gasch, A. P. *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–57 (2000).
31. Sarkisyan, K. S. *et al.* Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397–401 (2016).
32. Qian, W., Yang, J. R., Pearson, N. M., Maclean, C. & Zhang, J. Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.* **8**, (2012).
33. Brackley, C. A., Romano, M. C. & Thiel, M. The dynamics of supply and demand in mRNA translation. *PLoS Comput. Biol.* **7**, e1002203 (2011).
34. Sverdlov, A. V., Babenko, V. N., Rogozin, I. B. & Koonin, E. V. Preferential loss and gain of introns in 3' portions of genes suggests a reverse-transcription mechanism

- of intron insertion. *Gene* **338**, 85–91 (2004).
35. Cohen, N. E., Shen, R. & Carmel, L. The role of reverse transcriptase in intron gain and loss mechanisms. *Mol. Biol. Evol.* **29**, 179–86 (2012).
  36. Zhu, Y. O., Siegal, M. L., Hall, D. W. & Petrov, D. a. Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2310-8 (2014).
  37. Soemedi, R. *et al.* The effects of structure on pre-mRNA processing and stability. *Methods* **125**, 36–44 (2017).
  38. Gahura, O., Hammann, C., Valentová, A., Půta, F. & Folk, P. Secondary structure is required for 3' splice site recognition in yeast. *Nucleic Acids Res.* **39**, 9759–67 (2011).
  39. Warf, M. B. & Berglund, J. A. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.* **35**, 169–78 (2010).
  40. Munding, E. M., Shiue, L., Katzman, S., Donohue, J. P. & Ares, M. Competition between Pre-mRNAs for the Splicing Machinery Drives Global Regulation of Splicing. *Mol. Cell* **51**, 338–348 (2013).
  41. Man, O. & Pilpel, Y. Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species. *Nat. Genet.* **39**, 415–21 (2007).
  42. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**, 255–8 (2009).
  43. Voichek, Y., Bar-Ziv, R. & Barkai, N. Expression homeostasis during DNA replication. *Science* **351**, 1087–90 (2016).
  44. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
  45. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–4 (2011).



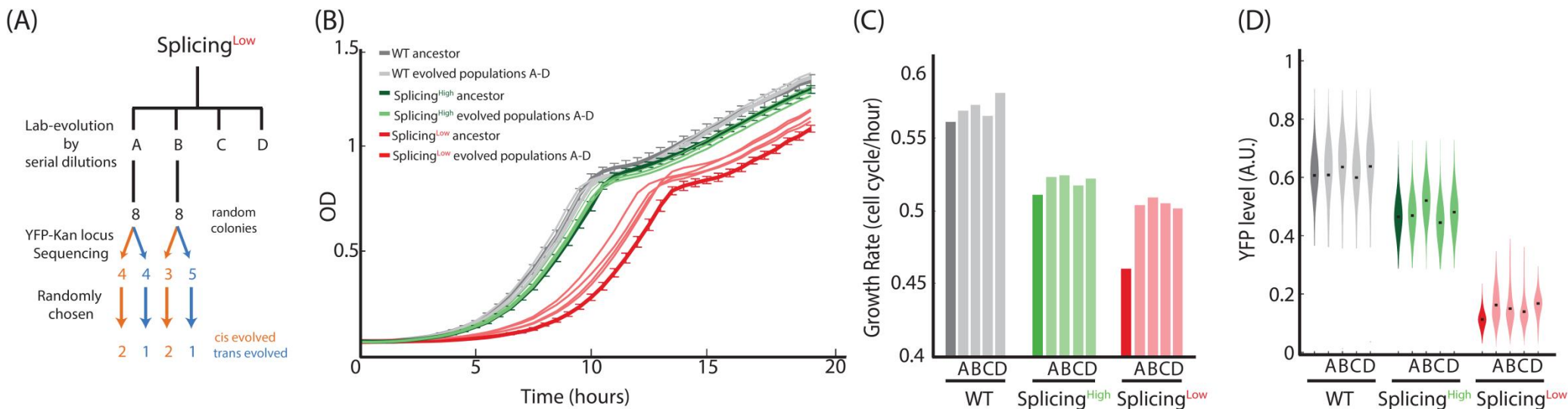
**Figure 1 – Inefficient intron splicing leads to lower gene expression levels and compromised antibiotics resistance.**

**A]** We introduced two alternative introns into a YFP domain that was fused to a kanamycin resistance domain - to generate three strains: (i) WT without an intron; (2) Splicing<sup>High</sup> with an efficiently spliced intron; and (iii) Splicing<sup>Low</sup> with an inefficiently spliced intron. Evolving cells at the presence of the antibiotics could adapt by mutating different parts of the YFP-Kan construct (evolution in *cis*) or other loci, evolution in *trans* (red stars represent potential locations of such putative mutation sites).

**B+C]** Splicing<sup>Low</sup> suffers from a severe growth defect compared to WT or Splicing<sup>High</sup> cells when the antibiotic is supplemented to the medium. The growth defect is manifested as a longer lag phase and a lower maximal growth rate.

**D]** Florescence intensity of the YFP-Kan reporter for all three strains shows that Splicing<sup>Low</sup> cells have lower expression levels of YFP-Kan. This observation links between YFP-Kan expression levels and cellular fitness.

**E]** Transcriptome profiling shows that ribosomal genes were down-regulated (green dots, p-Value=4.62x10<sup>-26</sup>, paired t-test) and stress-related genes were up-regulated (red dots, p-Value=3.40x10<sup>-5</sup>, paired t-test) in Splicing<sup>Low</sup> compared to WT cells. This observation suggests that Splicing<sup>Low</sup> cells are stressed because of compromised resistance to the antibiotics and that the general stress response was activated in them. **Inset]** Mean log<sub>2</sub> ratio of ribosomal and ESR gene groups.

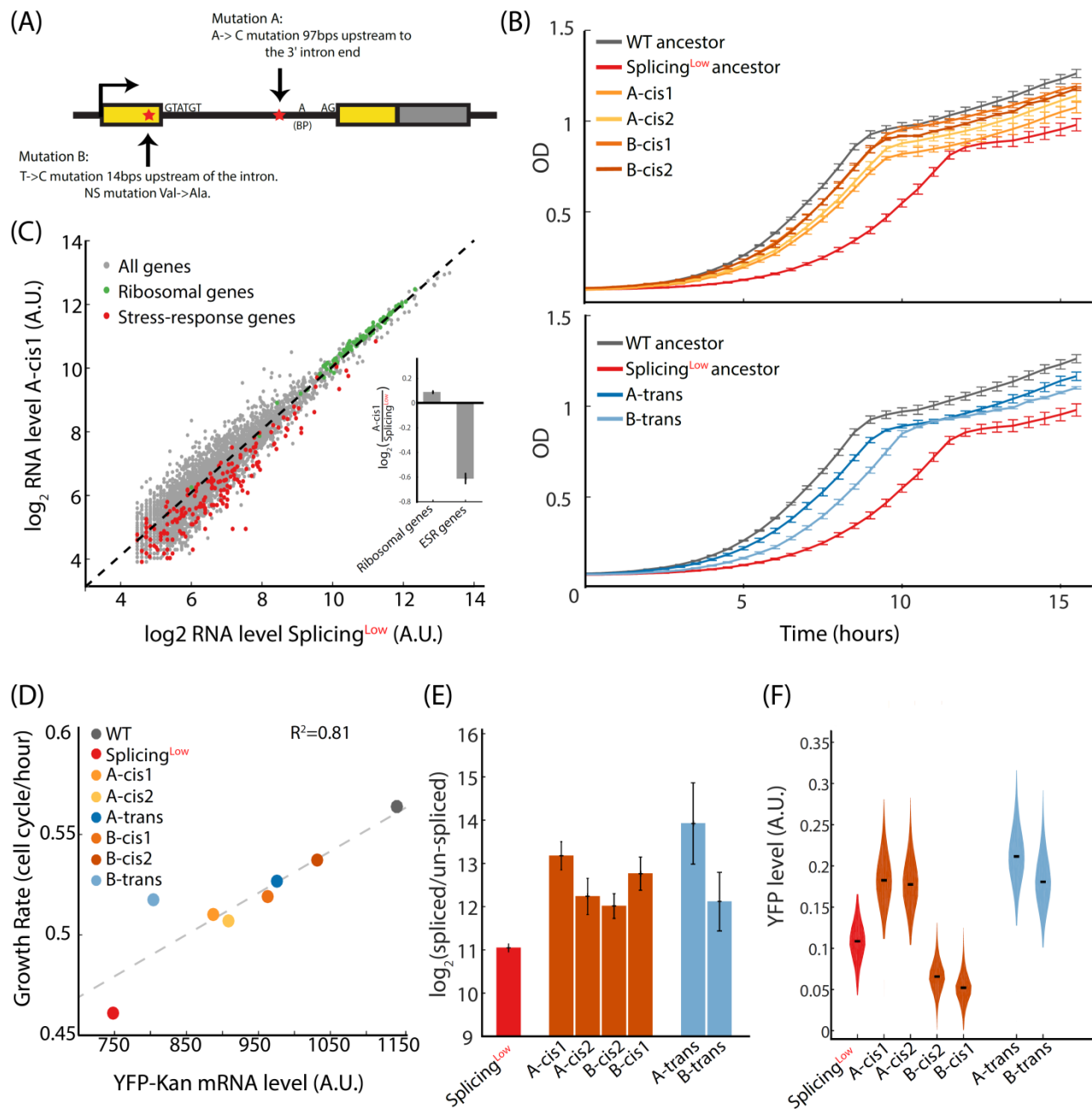


**Figure 2 – Rapid adaptation to the presence of the antibiotics is observed only for Splicing<sup>Low</sup> cells.**

**A|** We evolved WT, Splicing<sup>High</sup>, and Splicing<sup>Low</sup> cells for ~560 generations with the presence of the antibiotics in four independent cultures for each strain. We measured fitness and YFP-Kan expression levels for all evolved lines (see below), and also randomly chose 16 colonies from two evolved lines of Splicing<sup>Low</sup>. We sequenced the YFP-Kan locus of those colonies and observed that around half showed mutations in the YFP-Kan construct (indication of evolution in *cis*) and the other half did not (indication of evolution in *trans*). Of those colonies, we randomly chose two *cis*-evolved and one *trans*-evolved colonies from each evolved population for further examinations (see figure 3 onwards).

**B+C|** Growth of evolved populations compared to the three ancestors. Only evolved Splicing<sup>Low</sup> cells demonstrate significant improvement in growth for all four independent evolution lines. This observation suggests that the inefficiently spliced intron led to a rapid adaptation of Splicing<sup>Low</sup> cells.

**D|** Florescence intensity of the YFP-Kan reporter for all evolved cultures show that expression levels were much increased in all four evolved cultures of Splicing<sup>Low</sup> compared to the ancestral strain (effect sizes = 78.67, 79.54, 75.17, 83.19). Conversely, the increase in expression levels in the evolved WT and of Splicing<sup>High</sup> populations were smaller (WT effect sizes = 64.66, 68.44, 63.51, 67.74; Splicing<sup>High</sup> effect sizes = 54.33, 70.66, 52.43 and 58.27). This observation suggests that adaptation of Splicing<sup>Low</sup> cells was based on their ability to increase expression levels of the resistance proteins.



**Figure 3 – Evolved colonies demonstrate increased splicing efficiency that results in higher transcript levels and relieved stress.**

**A|** Sequencing of the YFP-Kan construct in the evolved colonies revealed two mutation types: (i) in the intron itself and (ii) in the up-stream exon – see text for full description. These mutations did not occur in the intron 5' donor, 3' acceptor, or the branching point – suggesting that other positions of the intron and its vicinity are phenotypically functional and may affect splicing efficiency.

**B|** All *cis*-evolved colonies (upper graph) and *trans*-evolved colonies (lower graph) show increased fitness compared to the Splicing<sup>Low</sup> ancestor, yet still lower than the WT ancestor.

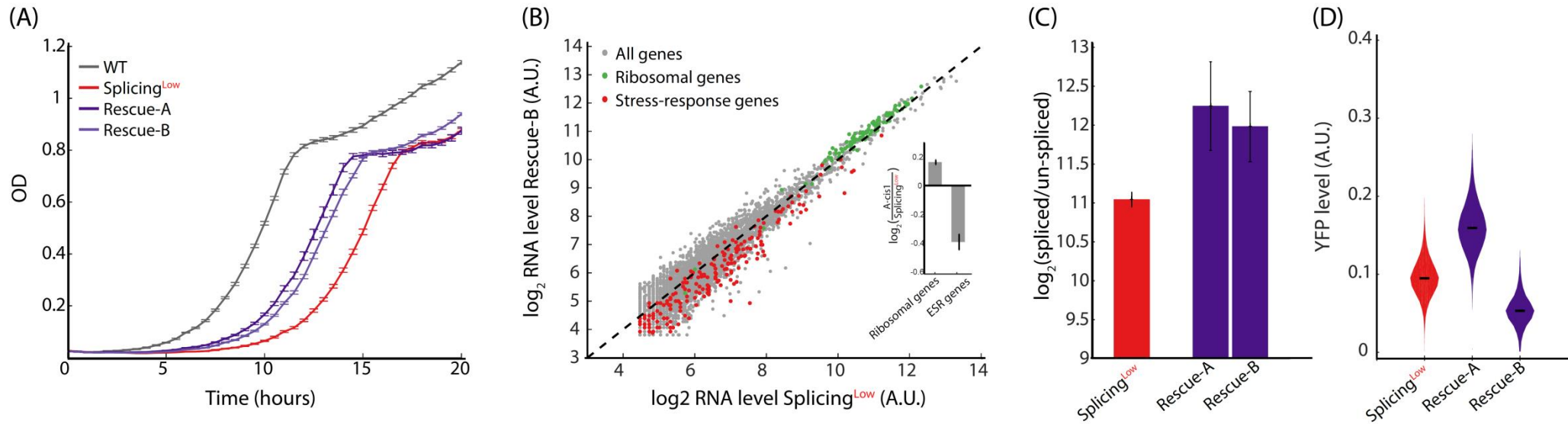
**C|** Transcriptome profiling reveals that ribosomal genes were up-regulated (green dots,  $p$ -Value= $4.94 \times 10^{-18}$ , paired t-test) and stress-related genes were down-regulated (red dots,  $p$ -Value= $3.64 \times 10^{-15}$ , paired t-test) in the evolved colony A-cis1 compared to the Splicing<sup>Low</sup> ancestor. This trend was observed in 5 out of 6 evolved colonies (Supplementary Figure 1). **Inset|** Mean  $\log_2$  ratio of ribosomal and ESR gene groups.

**D|** mRNA levels of YFP-Kan transcripts correlate with growth rate – suggesting that cellular fitness in our set-up is indeed determined by the availability of Kanamycin-resistance proteins to overcome the antibiotics.

**E|** All *cis*- and *trans*-evolved colonies demonstrate increased splicing efficiency of the YFP-Kan mRNA compared to the Splicing<sup>Low</sup> ancestor. This result suggests that all adaptation trajectories led to the adaptation of the splicing process to better mature the un-spliced YFP-Kan transcript.

**F|** Fluorescence intensity of the YFP-Kan reporter show increased levels for the two *cis*-evolved colonies with the mutation in the intron and for the two *trans*-evolved colonies. In contrast, the two *cis*-evolved colonies with the non-synonymous mutation in the exon demonstrate decreased YFP-Kan levels. This observation suggests that the non-synonymous mutation hampered the ability of the YFP domain to fluorescent and reduced the Fluorescence intensity per protein molecule (see text for full explanation).





**Figure 4 – cis-acting mutations are sufficient to increase fitness by elevating splicing efficiency.**

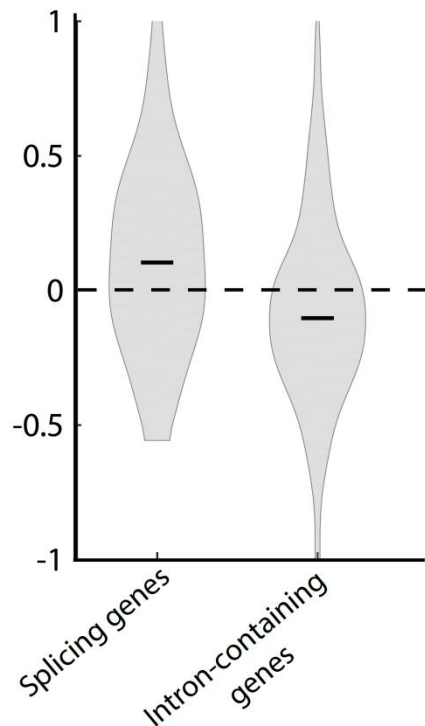
**A** | We created two rescue strains, each harboring one of the mutations that appeared spontaneously in the evolved populations. Growth of the two rescue strains show that a single mutation in the YFP-Kan construct is sufficient to increase fitness compared to Splicing<sup>Low</sup>.

**B** | The exonic mutation is also sufficient to alleviate stress, as ribosomal genes were up-regulated (green dots, p-Value= $1.02 \times 10^{-18}$ , paired t-test) and stress-related genes were down-regulated (red dots, p-Value= $9.02 \times 10^{-12}$ , paired t-test) in Rescue-B compared to Splicing<sup>Low</sup>. The same trend was also observed for the intronic mutation for Rescue-A cells. **Inset** | Mean  $\log_2$  ratio of ribosomal and ESR gene groups.

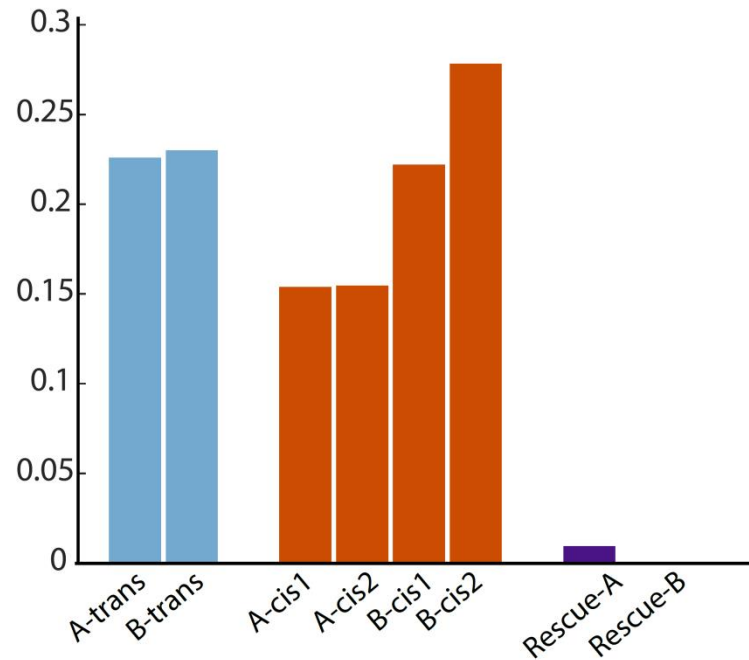
**C** | The two rescue strains demonstrate higher splicing efficiency of the YFP-Kan mRNA compared to the Splicing<sup>Low</sup> ancestor. This result suggests that a single mutation is sufficient to improve splicing efficiency.

**D** | Florescence intensity of the YFP-Kan reporter for the Rescue-A and Rescue-B strains show similar trends as the colonies in Figure 3D - supporting earlier conclusions.

(A) mRNA level  $\log_2 \left( \frac{\text{A-trans}}{\text{Splicing}^{\text{Low}}} \right)$



(B)  $\frac{\text{Supply}}{\text{Demand}} = \log_2 \left( \frac{\text{Fold change splicing genes}}{\text{Fold change intron-containing genes}} \right)$

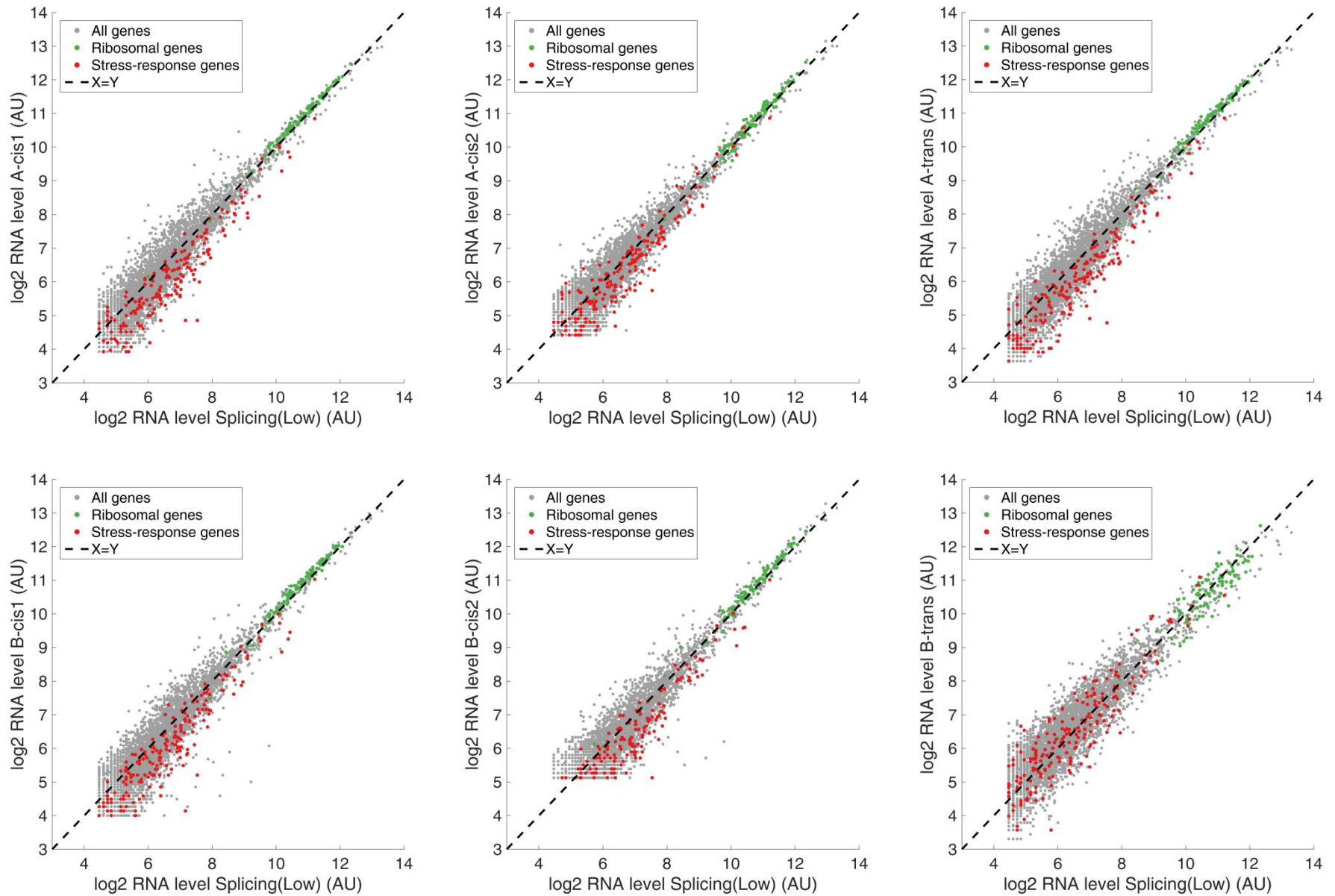


**Figure 5 – Increasing cellular availability of the splicing machinery is an adaptive mechanism of splicing.**

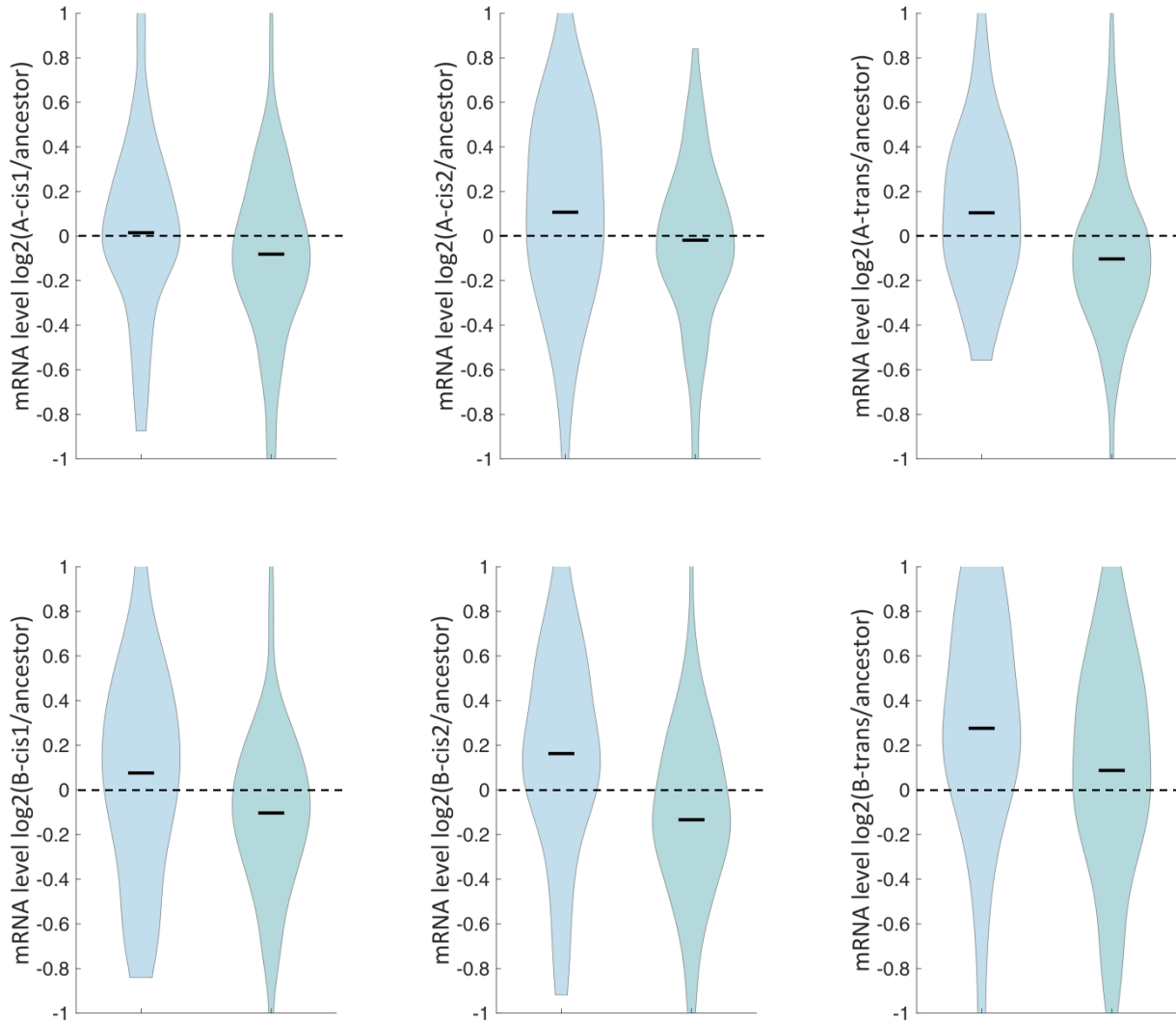
**A|** The groups of splicing genes and intron-containing genes were increased ( $p\text{-Value}=1.36 \times 10^{-3}$ , paired t-test) and decreased ( $p\text{-Value}=1.67 \times 10^{-2}$ , paired t-test), respectively, in the trans-evolved colony A-trans compared to  $\text{Splicing}^{\text{Low}}$  ancestor. This observation suggests that the supply-to-demand ratio of the splicing machinery was increased in A-trans colony, which allowed its increased splicing efficiency of the YFP-Kan transcript.

**B|** Supply-to-demand ratios for the splicing machinery were calculated to all evolved colonies and to the rescue strains as the difference between the mean fold-change of splicing genes to the mean fold-change of intron-containing genes. While supply-to-demand ratios were increased in all evolved colonies, they remained the same for the two rescue strains. These results suggest that indeed the cellular availability of the splicing machinery was elevated in the evolved colonies – a *trans*-adaptation mechanism to optimize gene expression using the splicing process.





**Supplementary Figure 1** | Transcriptome profiling reveals that ribosomal genes were up-regulated and that stress-related genes were down-regulated in 5 out of 6 of the evolved colonies compared to the Splicing<sup>Low</sup> ancestor.



**Supplementary Figure 2** | Supply-to-demand ratio in each of the evolved colonies. Folding change ratios in log2 are shown for splicing genes (left) and intron-containing genes (right). Black line represents the median of the distribution.