עבודת גמר (תזה) לתואר
מוסמך למדעים

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

מאת
עידן פרומקין

Thesis for the degree
Master of Science

Submitted to the Scientific Council of the
Weizmann Institute of Science
Rehovot, Israel

By
Idan Frumkin

על המנגנונים המניעים את האבולוציה של משפחות גנים מרובות עותקים של רנ״א שליח

What are the mechanisms driving the evolution of tRNA multi-copy gene families?

מנחה:
פרופ׳ יצחק פלפל

Advisor:
Prof. Yitzhak Pilpel

טבת ה'תשע"ג

December 2012

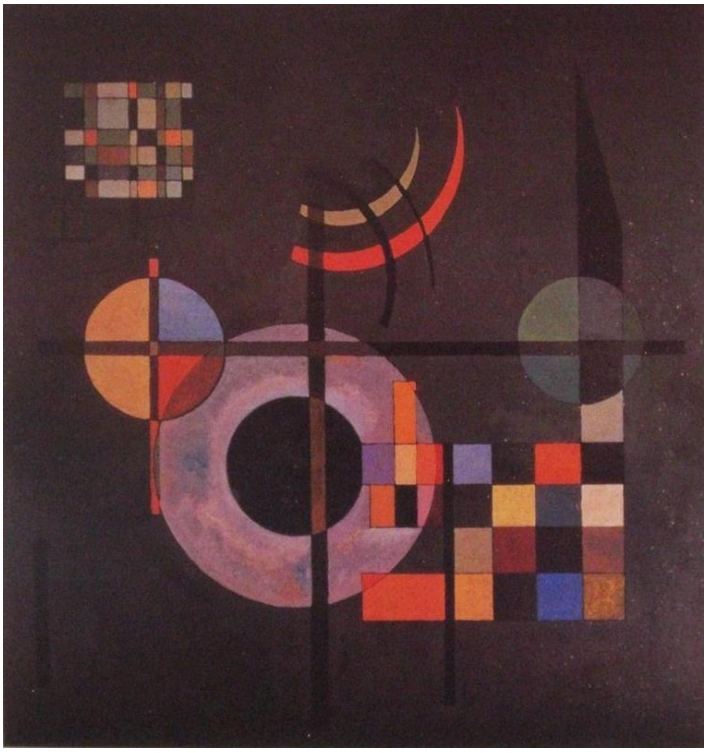
Acknowledgments

I would like to thank my mentor, Prof. Yitzhak Pilpel, for his continuous support and caring. His enthusiasm for biology makes the scientific process a more enjoyable playground than it already is. Thank you!

I wish to thank all my lab members for many fruitful discussions. You guys taught me of the endless ways science can be performed. Special thanks go to Orna Dahan, Avihu Yona & Zohar Bloom- you made this year full of stories and memories.

To Dr. Maya Schuldiner and her lab members – Tslil Ast, Yifat Cohen & Ido Yofe – I'd like to express my gratitude for their help and company since my very first step in science. I would also like to thank all my friends and peers for their tolerance and wisdom when giving me advice.

Finally, I will be forever grateful to my close family for their tremendous support and guidance throughout my path in life.



WASSILY KANDINSKY
Counter weights | 1926

1 | Abstract

Genomes often evolve via duplications of genetic material that may diverge over time. One result of such duplication events is the creation of multi-copy gene families, each comprised of all copies of a gene unique to that family. Though distinct, the gene copies in a family tend to be highly similar in sequence. Although much effort has been made to elucidate the evolution of duplicated genes, key questions concerning the evolution of multi-copy gene families are still elusive. We study multi-copy tRNA gene families in order to learn about the evolution of the tRNA pool. In the first part of my thesis, we focused on the 'translational balance' between the demand, the codon-usage, and the supply, the expression level of each tRNA type. To reveal the importance of the translational balance, we introduced a single-copy tRNA gene deletion in a yeast strain, which breaks the demand-to-supply balance. Lab-evolution experiments demonstrated that the balance was rapidly restored by mutations in other tRNA genes. Strikingly, parallel experiments showed that cells always found the same evolutionary solution: other tRNA genes, coding for the same amino acid, were mutated precisely in the third anti-codon position to become identical to that of the deleted tRNA. Yet, when we artificially added multiple copies of the single-copy tRNA to the wild-type, the cells exhibited proteotoxic stress suggesting an optimum in expression of this tRNA type. Hence, we show how tRNA genes change in evolution to meet translational demand on the one hand while maintaining a balanced gene family size on the other.

In the second part of my thesis, we focused on elucidating the mechanism behind the surprising, often perfect, sequence identity of tRNA multi-copy gene families. This conservation could be either the result of selective pressure to maintain a desired sequence or due to a 'concerted evolution' scenario that utilizes recurrent recombination events among family members. To this end, we developed a methodology based on a mutated copy of a tRNA gene which we cloned into high copy number plasmids and transformed into either a wild-type strain or strains carrying mutations in recombination pathways. Then, we allowed the strains to undergo a lab evolution process in tens of parallel cultures. At the end of the evolutionary period, we sequenced all genomic copies of the gene as well as the

dozens of mutated plasmid copies. We reasoned that if recombination among family members is responsible for preserving the perfect sequence identity, it will be visualized as homogenization of the populations of the endogenous and plasmid genes. Conversely, if selection is responsible for the maintenance of the high sequence identity, then the genomic sequence will be preserved. Since the genomic loci did not present any change in the tRNA sequence, our results support the possibility that the sequence identity is maintained by purifying selection. Alternatively, insufficient recombination rates in the current setup and time frame may have not been sufficient to contaminate the canonical genomic tRNA copies with mutations. Finally, we discuss the implication of our studies on some unresolved questions concerning the mechanisms driving the evolution of tRNA multi-copy gene families.

2| Introduction

Gene duplications, which occur in all kingdoms of life, play an important part in the evolution of genomes and is cardinal to the creation of biological innovation[1]. However, our current understanding of the evolutionary mechanisms involved in the emergence, maintenance and evolution of gene duplications is limited. One possible outcome of gene duplication events is the creation of many copies of the same gene in a single genome. The evolutionary process of such multi-copy gene families is termed “concerted evolution”, a process that maintains high sequence similarity among the copies and is balanced by selection, mutation, recombination and genetic drift[2–8].

Multi-copy gene families are subjected to genomic changes, such as mutations, that may results in divergence of the copies into paralogous genes with different functions[1]. On the other hand, sequence divergence between copies of a gene family may be low because of frequent exchanges of genetic information among the copies. The homogenization of sequences is achieved via various homologous recombination pathways and mostly by gene conversion, during which the DNA sequence of the donor allele is copied to the recipient loci while the recipient allele is lost[7,9–14]. The duration of concerted evolution is the time the copies of the same gene undergo evolution in a dependent manner. This evolutionary process ends once the copies have sufficiently diversified to the point that they are no longer subject to gene conversion. Additionally, insertions or deletions may terminate concerted evolution because they might work as a barrier against the pairing of paralogs and suppress gene conversion. Thus, the length concerted evolution period depends primarily on selection, mutation and gene conversion rates[8,15].

Indeed, selection for divergence of copies can work as a mechanism to terminate concerted evolution. Suppose that a new mutation with a novel function occurs in one of the duplicated copies, while the other keeps its original sequence. If the neofunctionalization of the copy is favored by strong selection, it can be maintained even under the pressure of homogenization by gene conversion[6,15]. However, selective pressure can also work in the same direction as gene conversion and maintain a favorable sequence throughout all copies of a gene family[7].

The theoretical applications of concerted evolution was studied extensively in past years[2–8]. Usually, these works decipher models that deal with a two-locus situation with respect to fixation of new mutations or polymorphism equilibrium. These models demonstrate that the duration of concerted evolution is highly varied and reduced for low rates of gene conversion[8]. The fixation probability of a new mutation is classically known to depend on its selection coefficient and the population size, yet surprisingly the probability of fixation is independent of gene conversion rate. However, the fixation time is dependent on all these three parameters[7]. Furthermore, when gene conversion rate is high, the dominant factor to determine the fixation time is selection, and when it is low, both selection and gene conversion rates play the major roles[7]. In addition, the theoretical works predict that the rate of non-synonymous mutations would be faster in a multi-copy gene family, if many sites are subject to positive selection. On the other hand, the rate should be slower if it is under the pressure of strong negative selection[7].

Evidences for concerted evolution rely mostly on bioinformatics and comparative genomics. For example, the duration of concerted evolution for each of the ~450 pairs of duplicated genes in *S. cerevisiae* was estimated and shown to be varied. Moreover, the duration of concerted evolution is positively correlated with gene expression, which can be explained by a cellular need for a product level that is achieved by more than a single copy of the gene[5]. Concerted evolution was also bioinformatically studied in *Drosophila melanogaster*. Comparing genomic analysis of different strains revealed a number of duplicated blocks undergoing concerted evolution by gene conversion[16]. Yet, the decay of such blocks over time was observed raising the question of the importance of concerted evolution in the final outcome of the shaping of the genome.

Most of the direct experimental evidence, and some additional computational work, for concerted evolution were demonstrated on rRNA genes. Unlike common duplicated genes, rRNA genes, together with tRNAs, are a unique subset of genes that appear in the genome in a high copy number. This fact stems from the inability of non-coding genes to rely on translation to tune their expression levels[17]. To

overcome this, genes encoding for RNA products have often evolved to be expressed from many genomic loci and are suggested to undergo concerted evolution[18].

Ribosomal DNA (rDNA) is comprised of tandem repeats and is ubiquitous to all living organisms. In eukaryotes, each rDNA unit contains the genes for the ribosomal RNA (rRNA) that are interspersed by spacer regions. These spacers include two internal transcribed spacer regions (ITS), two external transcribed spacers (ETS) and an intergenic spacer (IGS)[19,20]. It was previously shown that repetitive genes, such as rDNA, tend to lose copies through homologous recombination[21]. Yet, a unique amplification system that functions to maintain a high copy number of the rDNA repeats was characterized in *S. cerevisia*[21]. In short, when the copy number is reduced by recombination, DNA double-strand breaks (DSBs) are induced and are repaired in a way that increases the rDNA repeat copy number by replication of other repeats[21]. Through this recombination-based copy number regulation mechanism, the sequence of the rDNA repeats may be conserved throughout all copies of a species.

Studies regarding the homogeneity of rDNA repeats gave opposing observations. At first, evidence for strong concerted evolution for rDNA was observed in hominoid[22], *Nicotiana* allotetraploids species[23], *Drosophila*[24], yeast[25–27] and bacteria[28]. However, recent whole-genome shotgun sequencing analyses on rDNA of 12 *Drosophila* and several fungi species revealed a higher diversity in the rDNA repeats than initially observed[29,30]. Concurrently, comparative analysis revealed that various *S. cerevisiae* strains possess different patterns of rDNA polymorphism, with much of the variation located within the rapidly evolving IGS region[29]. In addition, the level of intra-individual variation in ITSs of the meadow grasshopper *Chorthippus parallelus* proved higher than expected[19]. In contrast, a natural species hybridization event between *Flammulina velutipes* and *Flammulina rossica* resulted in a homogenized ribosomal repeat containing elements of both parents[31]. Furthermore, a homogenization bias toward the maternal copy of the ITS in a hybrid from two scallop species, *Chlamys farreri* and *Argopecten irradians*, was observed during early development of the hybrid[32]. Importantly, changes in the rDNA loci of *Drosophila melanogaster* were observed in respect to the number of rRNA repeats and variation

in the IGS after growth of 400 generations in the lab[33]. The same phenomenon was reported in a similar experiment in *Daphnia obtusa* that lasted 90 generations and was established from a single wild caught female[34].

Compared to the large amount of data established for rDNA, there are very few evidences for concerted evolution in tRNA genes. tRNAs differ from rRNA genes in at least 3 aspects. First, although tRNA families may be present in genomes in as many as 20 copies, their copy number in much smaller numbers than the rDNA repeats, which appear in as many as 200 copies[35]. Second, the length of a typical tRNA gene (without introns) is 72bps[36,37] and is much shorter than the ~9kb rRNA repeat. Since the efficiency of homologues recombination depends on the length of the sequence similarities between the loci[11,12,14], this argues against a recombination-based mechanism that would maintain the sequence similarity among tRNA multi-copy gene families. Lastly, while the rRNA genes are localized to a specific chromosome and form the nucleolus, tRNAs are scattered across a variety of chromosomes. On the one hand, this may reduce the probability for any two tRNA copies to find themselves in proximity, a requirement for homologous recombination[38]. On the other hand, in a recent 2010 paper, which inferred the 3D structure of the yeast genome, two co-localized tRNA clusters were identified. One, in the nucleolus, that seems to be co-localized with the rRNA genes on chromosome XII, and another that seems to be clustered with centromeres[39]. Concurrently, tRNAs were shown to co-localize with 5S ribosomal DNA at the nucleolus in a transcription-dependent manner[40]. Thus, although the tRNA genes are not in linear proximity within chromosomes, they might co-localize in the folded structure of the genome.

The possibility that tRNA genes undergo concerted evolution must require that they have a significant recombination potential. Indeed, recombination and gene conversion of tRNA genes were demonstrated in several studies[41]. First, a reporter gene was placed on a plasmid between two neighboring tRNA copies in *S. cerevisiae* and its frequency of loss was measured. Surprisingly, the recombination rate was elevated when the genes were transcriptionally active[42]. Second, a tRNA gene was placed on a plasmid between two direct repeat elements and mitotic recombination between the repeats was measured in *S. cerevisiae*[43]. The presence of a tRNA gene

resulted in a significant increase in inter-repeat recombination. Again, the increased level of recombination was dependent upon the tRNA gene being actively transcribed, which suggests that Pol III-dependent transcription increases recombination rates. In addition, mutation of the RRM3 gene, involved in removing potential barriers to DNA replication, also increased the tRNA-dependent inter-repeat plasmid recombination. Thus, it is proposed that active tRNA genes stimulate recombination due to collisions between the Pol III complex and the DNA replication machinery[43]. Furthermore, tRNA genes are associated with both chromosomal breakpoints[44] and general genomic instability[41] that may facilitate homologous recombination near their genomic loci.

In the early and mid- 1980's, it was demonstrated that genetic information can transfer between serine tRNA copies in *Schizosaccharomyces pombe*[45–47]. These findings rely on a gene conversion event between an inactive suppressor tRNA to one of its two active siblings. In such event, the mutated anti-codon is moved to an active tRNA copy and thus allows growth on selective media. Indeed, these studies demonstrate genetic exchange among tRNA genes in yeast and suggest that concerted evolution is possible for them. However, the experiments described above are performed in a situation of a strong selection pressure and may not mimic the natural environment in which tRNAs have evolved.

Thus, questions regarding concerted evolution in general, and for tRNA genes specifically, are still not resolved. How often and how prevalent is concerted evolution? How long does it last and what are the terms for it to occur? What is the evolutionary significance of this phenomenon[8,15,48]? Since concerted evolution was studied mostly with the respect of intra-chromosomal gene conversion for rDNA repeats, its importance and prevalence for inter-chromosomal multi-copy gene families, such as tRNAs, is still an open question.

In addition to the relevance of concerted evolution in the shaping the tRNA pool, many other question regarding the evolution of tRNA genes are still elusive. One interesting question revolves the balance in the cell between the supply and demand during translation elongation step of a nascent polypeptide.

The process of translation elongation is governed by three parameters: the mRNA secondary structure[49], the codon usage of the mRNA[50], and the cellular tRNA pool[51]. The rate limiting step of polypeptide synthesis is the recruitment of a correct tRNA from the cellular tRNA pool that matches the translated codon[51]. Each tRNA gene exists as a gene family with several copies scattered throughout the genome. Moreover, it has been experimentally observed for *S. cerevisiae*[52] and *E. coli*[53], that the *in vivo* concentration of tRNA molecules is correlated to the number of copies coding for this tRNA type[54,55]. In *S. cerevisiae*, the genomic code is decoded by 42 different tRNA families that are present in its genome in a total of 274 tRNA genes. Each tRNA family ranges from 1 to 16 copies, with 6 tRNA families consist only of one copy.

The translation efficiency of a coding sequence is commonly defined as the extent to which its codon usage is adapted to the cellular tRNA pool[56,57]. Codon usage bias, the non-random use of synonymous codons, enhances translation efficiency by improving both elongation rates and translation accuracy[58]. In organisms such as *S. cerevisiae*[59], *C. elegans*, *D. melanogaster* and *A. thaliana*[60], codon bias was found to be most extreme in highly expressed genes, for which "optimal" codons are frequently used[61]. The evolutionary force that drives the correlation between codon bias and expression levels was coined translational selection. This term refers to the natural selection acting to maintain high translation efficiency for highly expressed genes[57].

In light of translational selection, the tendency of organisms to maintain some of their tRNAs in a low copy number is unclear and results in "non-optimal" codons. Yet, the fact that this phenomenon is conserved throughout different organisms raises the possibility that the existence of certain tRNA in low copies is biologically beneficial. It has been proposed in several studies that "non-optimal" codons may play a role in governing the process of co-translational folding, by slowing the process of translation to facilitate proper folding[62–65]. Yet, the contribution of non-optimal codons to proper protein folding was observed only for individual genes. Furthermore, the extent and importance of this phenomenon to the global folding state of the proteome remains elusive.

Thus, there is a balance between supply and demand for tRNAs in the cell that is significant to maintain a proper protein production. Despite the importance of this delicate balance, little attention has been paid to the mechanisms that govern the evolution of the supply. Under which conditions will the tRNA genetic network be ultimately changed?[66]. Specifically, it is unclear whether and how can the cell adapt in an evolutionary time scale to changes in the environment that break the delicate balance between tRNA supply and codon demand.

3 | Goals

The aim of this thesis is to elucidate the mechanisms that drive the evolution of tRNA multi-copy gene families. We decided to ask so far un-tackled questions regarding the evolution of the supply, namely the tRNA pool. We also aimed to understand the place of a concerted evolution process in the shaping of the tRNA cellular pool. Particularly, we aimed to elucidate how tRNA multi-copy gene families can maintain high degree of sequence identity among themselves.

Goal I

In the first part of this thesis, we focused on the translational demand vs. supply balance and explored its significance to the integrity of the proteome in yeast. Our goal was to mutate a tRNA gene, which results in a break of the translational balance, and understand the pathways cells can take toward evolutionary adaptation to this genetic challenge.

Goal II

The goal of the second part of my thesis was to elucidate whether tRNA genes evolve dependently on one another in a concerted manner. We aimed to elucidate whether the sequence conservation among the 11 copies of a tRNA multi- copy gene family is maintain via selection or due to neutrality, by gene conversion events. To do so, we introduced the cells with many copies of a mutated tRNA sequence and monitored the sequence of the other tRNA loci following lab-evolution experiments.

4 | Methods

Strains & Growth

All Saccharomyces cerevisiae strains were created on the genetic background of BY4741 (MATa his3Δ1, leu2Δ0, met15Δ0, ura3Δ0). The creation of the strains Δ*tR(UCU)*, Δ*rad51* and Δ*rad52* was performed via transformation of a G418 resistance marker to the genes' loci by homologous recombination. During all experiments, cells were grown at 30^{o}c in either YPD (rich medium) or SD (poor medium). When needed, Galactose was supplemented to the medium to induce expression of a gene. YPD or SD-Ura were used for the lab-evolution experiments.

Plasmid construction & Restriction Free Cloning

All plasmids were cloned using the Restriction-Free cloning methodology[67]. The gene tR(UCU) was cloned into a high copy 2μ pRS426 plasmid[68] with a ranging number of up- and down-stream bps. See text for full explanation on the plasmid construction.

Lab-Evolution Experiments

All laboratory evolution experiments were carried out by serial dilution. The lab-evolution experiments were performed in either 96well plate (each well contained 150μl of culture) or 24well plate format (each well contained 1.2ml of culture). Cells were grown until reaching stationary phase under the relevant condition and were diluted by a factor of either 1:75 (96well plate format, 6.5 generations per dilution) or 1:128 (24well plate format, 7 generations per dilution) into fresh media once a day.

Liquid Growth Measurements

Deep stationary (48h) cultures were diluted into fresh medium and were grown at the relevant condition. OD_{600}/YFP measurements were taken every 45mins until reaching stationary phase. Qualitative growth comparisons were performed using 96well plates in which two strains were divided on the plate in a checkerboard manner to cancel out geographical effects. For each strain, a growth curve was obtained by averaging over 48 wells.

Measurements for Saturation of the Protein Quality Control Machinery

We utilized a previously published method that allows examination of the protein quality control of the cell[69]. This assay provides an indication for the protein unfolding stress in cells by assessing the load on the protein quality control machinery. In this assay, cells were introduced with a high copy number plasmid that contains the human gene von Hippel-Lindau (VHL) fused to a florescent tag (mCherry). VHL is a naturally unstructured protein whose proper folding in human cells depends on a complex formation with two other proteins, Elongin B and C. Expressing VHL in yeast cells which lack VHL's complex partners, leads to misfolding of the translated proteins. Under normal conditions, the misfolded VHL proteins are handled by the cell's quality control machinery. When the quality control machinery is not saturated, the fluorescently tagged VHL appears in the cytosol. However, under stress, in which the quality control machinery is fully occupied, misfolded proteins in the cytosol are processed into dedicated inclusions (JUNQ and IPOD) and form punctum structures. Hence, a punctum phenotype of the VHL-mCherry construct is an indication that cells suffer from high protein misfolding levels and saturated protein quality control machinery. Selected strains were transformed with the plasmid and selected on SD-URA plates. Overnight cultures, were diluted (1:10) and visualized in the mid-log growth phase using an Olympus IX71 microscope controlled by Delta Vision software. Next, hundreds of cells (~1000) from each strain were counted for the number of cells with the puncta-phenotype and the results were normalized to the control strain.

5 | Results

5.1 tRNA Genes Rapidly Change in Evolution to Meet Translational Demand* [1]

Deletion of singleton tRNA gene breaks the translational balance

To demonstrate the importance of the balance between codon-usage and the cellular tRNA pool we created a yeast strain in which the single copy of an arginine tRNA gene, *tR(CCU)J*, was deleted (termed $\Delta tRNA^{Arg}_{CCU}$). Thus, the deleted cells cannot translate the arginine codon AGG with a fully-matched tRNA and need to relay on wobble interactions. This shortage in supply cannot match the demand for translation of AGG codons, which encompasses ~21% of the arginine codons in the yeast genome. Indeed, the $\Delta tRNA^{Arg}_{CCU}$ strain showed a severe growth defect compared to the wild-type strain (Fig. 1) indicating the major perturbation to the tRNA pool.

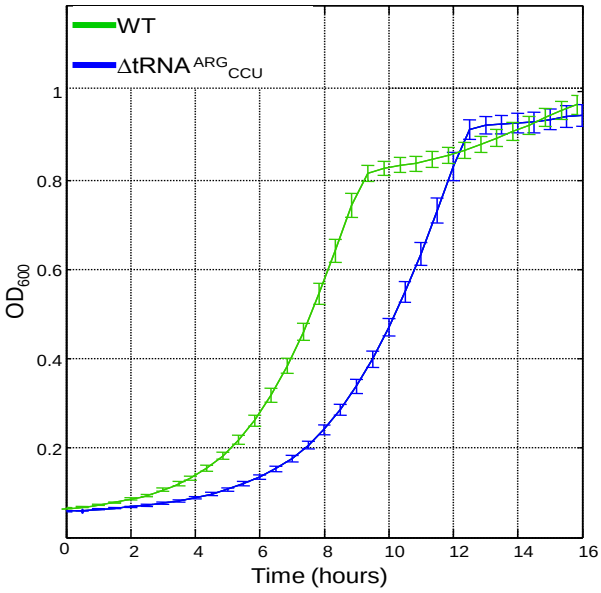


Figure 1 – The growth defect associated with deletion of the singleton $tRNA^{Arg}_{CCU}$
Growth curve measurements of WT (green) and $\Delta tRNA^{Arg}_{CCU}$ (blue) are shown in optical density (OD) values over time during continuous growth at 30°C.

* The following part was performed in collaboration with two students at the lab, Zohar Bloom & Avihu Yona.

Notably, the viability of ΔtRNA$^{Arg}_{CCU}$ cells is attributed to wobble base-pairing between the AGG codon and another arginine tRNA, tRNA$^{Arg}_{UCU}$. The severe growth defect of ΔtRNA$^{Arg}_{CCU}$ demonstrates that the capacity to efficiently translate AGG codons via wobble base-pairing is limited. Thus, the severe growth defect demonstrated by ΔtRNA$^{Arg}_{CCU}$ is most likely a result of the imbalance between the demand to translate AGG codons and the supply shortage of the altered tRNA pool caused by the deletion. Additional indications for the importance of the balance between the codon usage and the tRNA pool come from the other five singleton tRNA genes, as four of them are inviable upon deletion[70–72].

tRNA pool can rapidly evolve to meet cellular translational demands

In order to learn how cells can evolutionarily adapt to the translational imbalance in the tRNA pool, we performed a lab-evolution experiment via serial dilutions[73]. For that purpose, the ΔtRNA$^{Arg}_{CCU}$ strain was grown under optimal conditions and diluted into fresh medium by a factor of 120, equal to ~7 generations per day. Every 50 generations, growth of the evolving population was compared to both wild-type and the ancestor ΔtRNA$^{Arg}_{CCU}$ strain. Strikingly, after 200 generations a full recovery of ΔtRNA$^{Arg}_{CCU}$ growth defect was observed, i.e. the growth of the evolved population was indistinguishable from that of the wild-type (Fig. 2). It is worth noting that we evolved a WT strain in parallel to the lab-evolution of ΔtRNA$^{Arg}_{CCU}$ cells for the same time. Reassuringly, the growth of the WT strain did not improve in a significant manner, supporting the hypothesis that cells are able to adapt to a break in the translational balance

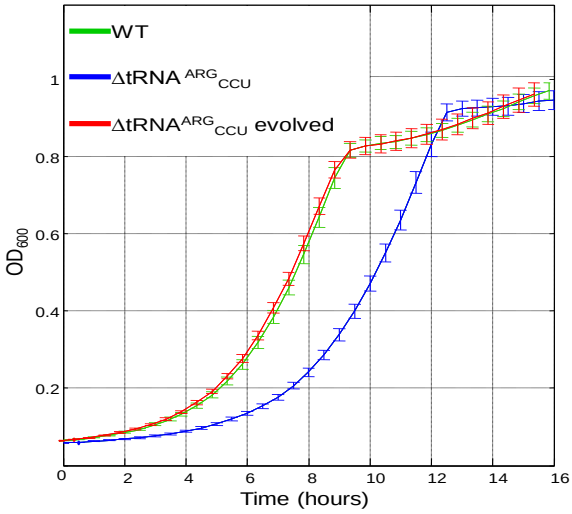


Figure 2 – Short lab-evolution cured the deletion phenotype of ΔtRNA$^{Arg}_{CCU}$

Growth curve measurements of WT (green), ΔtRNA$^{Arg}_{CCU}$ (blue) and the evolved deletion (red) are shown in OD values over time during continuous growth at 30°C.

A similar dynamics was observed in all four independent evolutionary lines of the experiment. In search for the genetic adaptations underlying this rapid recovery, we first looked for genetic alterations in other arginine tRNAs. We found a single point mutation in the arginine tRNA gene that codes for tRNA$^{Arg}_{UCU}$. This mutation changed the anticodon triplet of tRNA$^{Arg}_{UCU}$ from UCU to CCU (i.e. U→C transition). Consequently, the evolved tRNA$^{Arg}_{UCU}$ matches perfectly the AGG codon (Fig. 3).

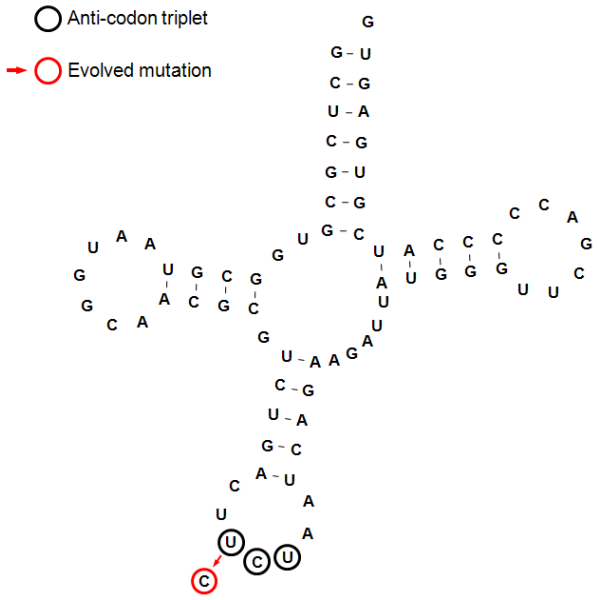


Figure 3 – During the lab-evolution of ΔtRNA$^{Arg}_{CCU}$, tRNA$^{Arg}_{UCU}$ mutated its anticodon to CCU
The secondary structure of tRNA$^{Arg}_{UCU}$ is drawn with the UCU anti-codon nucleotides marked with black circles. The red circle indicates the mutation that occurred during the lab-evolution experiments, i.e. U→C transition.

Unlike the singleton tRNA$^{Arg}_{CCU}$, there are 11 copies of tRNA$^{Arg}_{UCU}$ in the yeast genome. Although each of the four independent lab-evolution experiments showed the exact same solution, three different copies of this gene were changed in the four repetitions (i.e. one of the 11 copies was mutated in two of the repetitions). In order to confirm that the single point mutation in the anticodon of tRNA$^{Arg}_{UCU}$ is sufficient to eliminate the growth defect of ΔtRNA$^{Arg}_{CCU}$, we artificially inserted the U→C mutation into a ΔtRNA$^{Arg}_{CCU}$ background. The mutation was inserted into one of the 11 copies of the tRNA$^{Arg}_{UCU}$ genes, a copy that is located on chromosome XI that was spontaneously mutated in one of the lab-evolution lines. Indeed, the artificially mutated strain, termed here *MutΔtRNA$^{Arg}_{UCU}$*, showed full recovery of the deletion adverse phenotype (Fig. 4). This indicates that although additional mutations may have fixed in the genome of the evolved strains, the U→C mutation in the anticodon of tRNA$^{Arg}_{UCU}$ is sufficient for the full recovery of the tRNA$^{Arg}_{CCU}$ deletion phenotype.

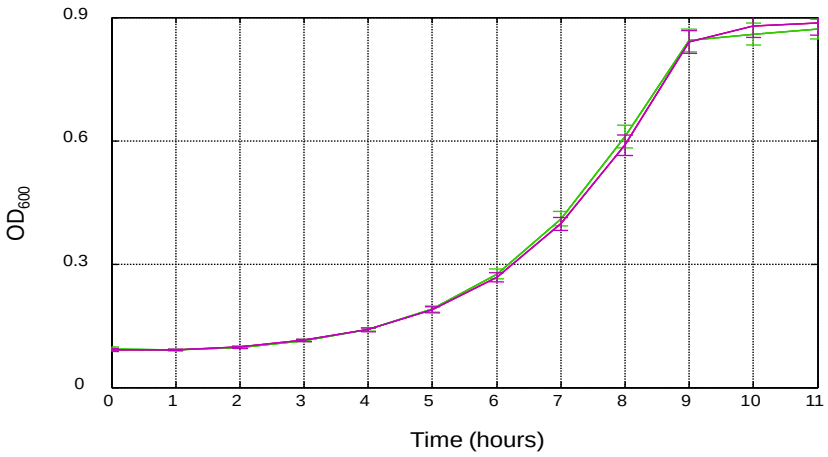


Figure 4 – *MutΔtRNA$^{Arg}_{UCU}$* exhibits similar growth as WT
A rescue experiment in which the U→C mutation was introduced to ΔtRNA$^{Arg}_{CCU}$ cells (termed *MutΔtRNA$^{Arg}_{UCU}$*) was performed. Growth curve measurements of WT (green) and of *MutΔtRNA$^{Arg}_{UCU}$* (magenta) are shown in OD values over time during continuous growth at 30°C.

Mutated tRNA$^{Arg}_{UCU}$ is fully functional despite sequence dissimilarities with respect to the deleted tRNA$^{Arg}_{CCU}$

The two arginine tRNA genes tRNA$^{Arg}_{UCU}$ and tRNA$^{Arg}_{CCU}$ differ in 21 out of 72 nucleotide positions, including the third anticodon position. Thus, the evolutionary solution that occurred in our experiments created a "hybrid" tRNA with a CCU anticodon while the rest of the tRNA sequence (termed here as the "tRNA body") remained as tRNA$^{Arg}_{UCU}$. Therefore, the body of the new hybrid tRNA, with the CCU anticodon, differs in 20 positions from the original tRNA$^{Arg}_{CCU}$ (Fig. 5).

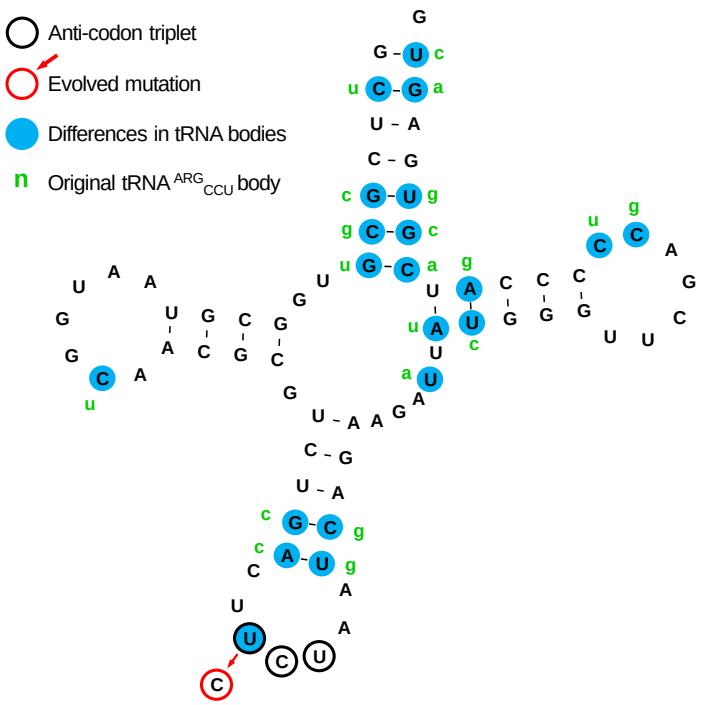


Figure 5 – Overview of the hybrid tRNA sequence created during the lab-evolution
The secondary structure of the hybrid tRNA is drawn. This hybrid is composed of the body of $tRNA^{Arg}_{UCU}$ with a CCU anti-codon. The anti-codon triplet is marked with black circles. The evolved mutation is marked with a red circle. All 20 nucleotide differences between $tRNA^{Arg}_{UCU}$ and $tRNA^{Arg}_{CCU}$ are marked with blue circles, next to which- in green letters- the nucleotide of $tRNA^{Arg}_{CCU}$ is written.

In general, the copies of a tRNA gene family tend to be highly similar in sequence[74], and in particular, the sequence of the 11 copies of $tRNA^{Arg}_{UCU}$ are 100% identical. A sequence similarity of that extent could suggest that the unique sequence of each tRNA body has a functional role in addition to that of the anticodon[75]. Therefore, despite these sequence differences, it is surprising that the hybrid tRNA performed just as well as the deleted $tRNA^{Arg}_{CCU}$ in terms of its effect on the measured growth of the strains. Thus, we raised the hypothesis that more challenging growth conditions might expose an inferiority of the hybrid tRNA, as was demonstrated more generally in the yeast deletion library[76,77]. To test this notion, we compared the growth of the rescued strain *Mut*Δ$tRNA^{Arg}_{CCU}$, which carries the hybrid tRNA, to that of the wild-type, under an array of unfavorable conditions. Surprisingly, under all checked conditions, there was no significant growth difference between the two strains (Fig. 6). Therefore, these results cannot support the hypothesis that the 20

nucleotide differences between the two tRNA bodies have a functional role that substantially affects growth. Hence, the hybrid tRNA provides an *in vivo* indication that the bodies of tRNAs that code for the same amino-acid are interchangeable despite extensive differences in their sequence. These results raised an alternative hypothesis (see section 2 of the Results) that the sequence identity among all copies of $tRNA^{Arg}_{UCU}$ is neutral and is a result of massive gene conversion events among the tRNA genes.

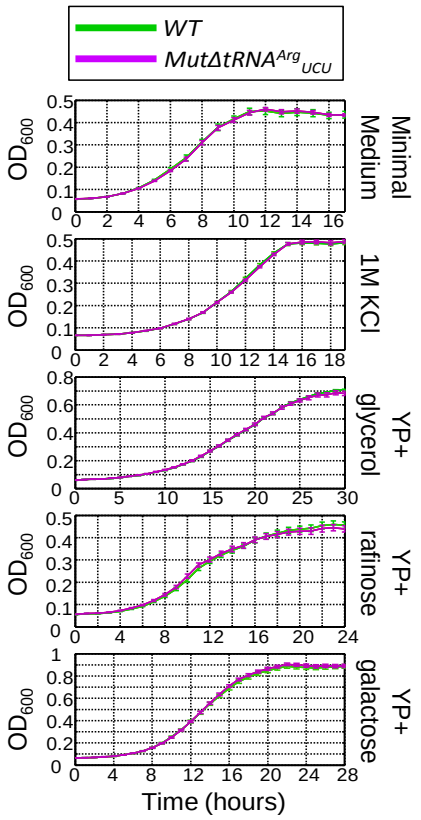


Figure 6 – The growth of *MutΔtRNA*$^{Arg}_{UCU}$ compared to WT under different conditions
Growth curve measurements of WT (green) and of *MutΔtRNA*$^{Arg}_{UCU}$ (magenta) are shown in OD values over time during continuous growth. Under all stresses tested (minimal medium, 1M KCl, YP+glycerol, TP+rafinose and YP+galactose) the rescued strain demonstrated similar growth as the WT strain.

To examine the generality of these observations, we again perturbed the tRNA pool in a wild-type strain by deletion of an entire serine tRNA family. Here, the supply of $tRNA^{Ser}_{GCU}$ was eliminated by deletion of all four identical genes of this family located on chromosomes *IV, VI, XII* and *XV*. A complete deletion of this gene family is lethal, indicating that the $tRNA^{Ser}_{GCU}$ is essential in *S. cerevisiae*. To validate that the

lethality is indeed due to the $tRNA^{Ser}_{GCU}$ genes deletion and not due to an unintentional perturbation of other putative genetic features in the vicinity of the deleted $tRNA^{Ser}_{GCU}$ genes, we introduced a plasmid with the $tRNA^{Ser}_{GCU}$ gene. Indeed, the deletion strain was viable in the presence of a plasmid carrying this gene. Then, we hypothesized that similarly to $tRNA^{Arg}_{CCU}$, other hybrid serine tRNAs that carry a GCU anticodon can prevent the observed lethality regardless of their tRNA body sequence. Therefore, rather than inserting a plasmid with the original $tRNA^{Ser}_{GCU}$ gene, we complemented the $tRNA^{Ser}_{GCU}$ family deletion strain with a plasmid containing a hybrid serine tRNA with GCU anticodon and two alternative bodies. The first was the body of $tRNA^{Ser}_{AGA}$ that differs in 23 positions from the deleted serine tRNA family and is found in the genome in 11 copies. The other was the singleton $tRNA^{Ser}_{CGA}$ that differs in 22 positions. Notably, the tRNA bodies of $tRNA^{Ser}_{AGA}$ and $tRNA^{Ser}_{CGA}$ differ in 12 positions between themselves. Indeed, both hybrid strains are viable on the background of the $tRNA^{Ser}_{GCU}$ family deletion. Therefore, we conclude that the identity of the anticodon is essential for the function of the $tRNA^{Ser}_{GCU}$ gene family. Yet, this essential function can be performed with hybrid serine tRNAs as long as the anticodon is GCU.

$tRNA^{Arg}_{CCU}$ is kept as a rare tRNA gene in various yeast species

Since our lab-evolution experiments showed a UCU→CCU transition in several arginine genes within only 200 generations, we asked what prevents such transition from naturally occurring in the genome throughout evolution in the wild-type, i.e. when the genome contains one copy of the $tRNA^{Arg}_{CCU}$ gene. Towards this end, we checked whether $tRNA^{Arg}_{CCU}$ is maintained at a low gene copy number also in other *S. cerevisiae* strains as well as among other yeast species. As a control we also compared the copy number of $tRNA^{Arg}_{UCU}$. We found a strong tendency to maintain $tRNA^{Arg}_{CCU}$ at a low-copy number. In all 27 *S. cerevisiae* strains checked[78,79], the $tRNA^{Arg}_{CCU}$ gene is found in a single copy. In comparison, the $tRNA^{Arg}_{UCU}$ gene is maintained at a high copy number, ranging from 8 to 12 copies. Concurrently, the tendency to maintain $tRNA^{Arg}_{CCU}$ as a single copy was also observed in other yeast species with the exception of *Aspergillus nidulans* that carries two copies of both

$tRNA^{Arg}_{CCU}$ and $tRNA^{Arg}_{UCU}$. In this yeast, unlike the other species, the codon usage of the AGG codon was very similar to that of the AGA codon (Fig. 7).

Yeast Specie	# of tRNAs with CCU anticodon	AGG usage in ARG codons (%)	# of tRNAs with UCU anticodon	AGA usage in ARG codons (%)
Saccharomyces cerevisiae	1	21	11	48
Candida albicans	1	8	5	56
Candida glabrata	1	20	9	52
Aspergillus nidulans	2	11	2	13
Debaryomyces hansenii	1	17	10	56
Encephalitozoon cuniculi	1	43	1	36
Eremothecium gossypii	1	14	6	16
Kluyveromyces lactis	1	18	7	51
Schizosaccharomyces pombe	1	11	2	23
Yarrowia lipolytica	1	5	4	17

Figure 7 – Various yeast species tend to keep $tRNA^{Arg}_{CCU}$ in a single copy

All examined yeast species, except for *Aspergillus nidulans*, keep a single copy of $tRNA^{Arg}_{CCU}$ compared to $tRNA^{Arg}_{UCU}$, which is mostly found in multiple copies.

Carrying multiple copies of the rare $tRNA^{Arg}_{CCU}$ gene is deleterious to the cell

The tendency of different yeast species, as well as different *S. cerevisiae* strains, to maintain $tRNA^{Arg}_{CCU}$ at a low copy number implies that selection drives the "natural state" of *S. cerevisiae*, in which $tRNA^{Arg}_{CCU}$ is found in a single copy. Such translational selection force can prevent the transformation of UCU anticodons to CCUs in arginine tRNA genes. Indeed, introduction of a multi-copy plasmid containing a $tRNA^{Arg}_{CCU}$ gene to a wild-type strain (termed here *WTmultiCCU*) resulted in a substantial growth reduction compared to wild type cells carrying a similar multi-copy plasmid that does not contain any tRNA gene (termed *WTmultiControl*). To control for the effect of multiple copies of tRNA genes in general, we also created a strain with a similar multi-copy plasmid that contains $tRNA^{Arg}_{UCU}$, termed here *WTmultiUCU*. Comparing the growth of *WTmultiCCU* and *WTmultiUCU* revealed that

WTmultiCCU has significantly reduced growth compared to *WTmultiUCU*, which exhibit a growth profile much closer to the profile of *WTmultiControl* (Fig. 8). These data are in line with the evolutionary tendency to keep low-copy number of $tRNA^{Arg}_{CCU}$ and suggest that high copy number of such rare tRNA genes is deleterious to the cell. Comparing the growth of wild-type cells transformed with centromeric-plasmid (low copy number plasmid) carrying either $tRNA^{Arg}_{CCU}$ or $tRNA^{Arg}_{CCU}$ showed a modest growth defect of the cells with the $tRNA^{Arg}_{CCU}$ plasmid, yet only under heat of 39°C (not shown).

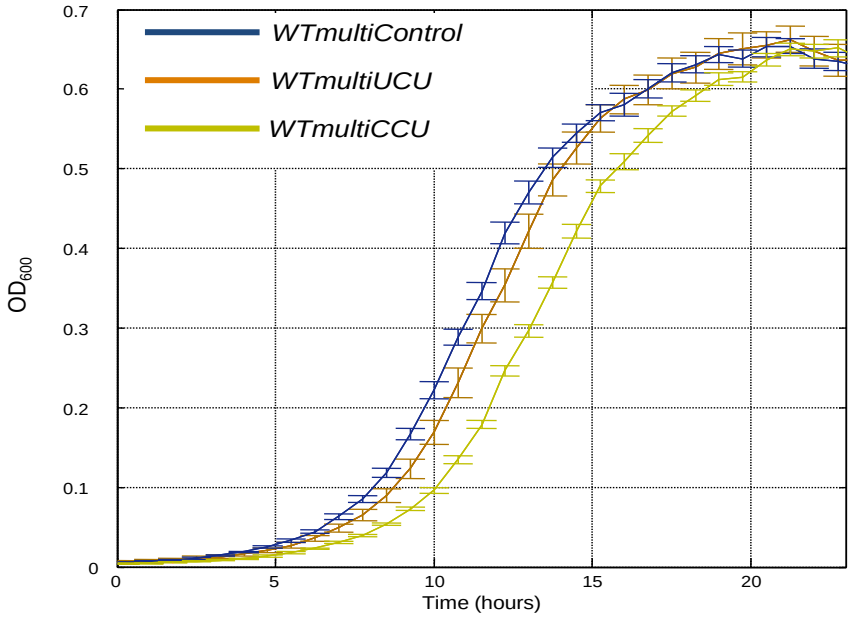


Figure 8 – *WTmultiCCU* suffers from growth defect compared to *WTmultiUCU*
Growth curve measurements of *WTmultiControl* (blue), *WTmultiUCU* (brown) and *WTmultiCCU* (khaki) are shown in OD values over time during continuous growth. The *WTmultiCCU* strain with a high copy number plasmid harboring $tRNA^{Arg}_{CCU}$ demonstrates a slower growth compared to cells with an empty plasmid or with $tRNA^{Arg}_{UCU}$.

Multiple copies of the rare $tRNA^{Arg}_{CCU}$ induce proteotoxic stress

There are several potential mechanisms that can cause the growth defect exhibited by *WTmultiCCU*: 1) Low concentration of certain tRNAs can be essential for proper folding of specific domains during the synthesis of a new protein[62,64,80,81]. 2) Mis-incorporation of arginine into non-arginine codons that may cause folding problems. 3) Mis-loading of arginine tRNA molecules with other amino acids that results in a wrong translation of the genetic code. While these potential errors are not mutually

exclusive and can each contribute to the observed growth defect, they all directly affect protein folding and can induce proteotoxic stress. To examine the possibility that the growth defect associated with multiple copies of $tRNA^{Arg}_{CCU}$ is indeed associated with proteotoxic stress, we utilized a previously published method that allows examination of the load on the protein quality control machinery (see methods)[69]. In this assay, we introduced cells with a plasmid that contains the human gene von-Hippel-Lindau (VHL) fused to a florescent tag (mCherry). When the fluorescently tagged VHL appears as a punctum phenotype (Fig. 9A), and not in the cytosol (Fig. 9B), this indicates high levels of misfolded proteins and a saturated protein quality control machinery. We transformed the VHL-mCherry plasmid to each of the multi-copy tRNA strains, *WTmultiCCU*, *WTmultiUCU* and *WTmultiControl* and monitored the level of proteotoxic stress by quantifying the number of puncta in each population. The fold change in puncta number was then deduced by normalization to the *WTmultiControl* population. We found that while *WTmultiUCU* exhibited similar puncta levels as the *WTmultiControl*, the *WTmultiCCU* exhibited a 3-fold increase in puncta levels (Fig. 9C). These findings suggest that high copy number of the rare tRNA gene, $tRNA^{Arg}_{CCU}$, but not an already abundant tRNA, $tRNA^{Arg}_{CCU}$, leads to elevated levels of proteotoxic stress.

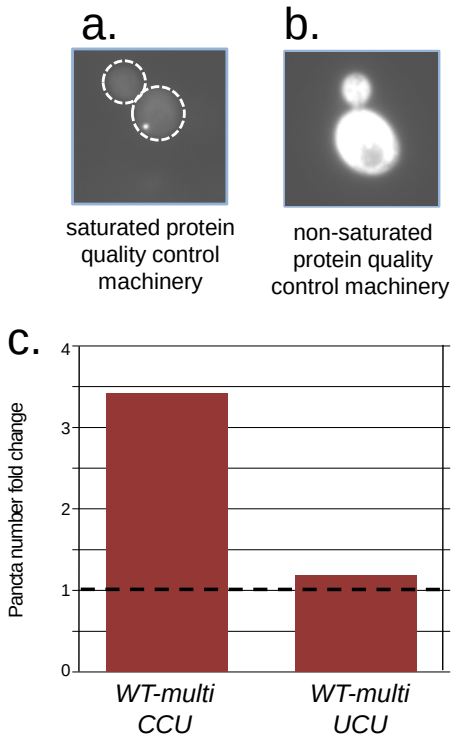


Figure 9 – *WTmultiCCU* demonstrates higher levels of misfolded proteins compared to *WTmultiUCU* **(a)** A *WTmultiCCU* cell in which the mCherry-VHL proteins appear with a punctum phenotype when the protein quality control machinery is saturated with misfolded proteins. **(b)** A *WTmultiUCU* cell in which the protein quality control machinery is not occupied with other proteins and mCherry-VHL is localized to the cytosol. **(c)** *WTmultiCCU*, *WTmultiUCU* and *WTmultiControl* were transformed with a mCherry coding plasmid and visualized under the microscope. 1000 cells per strain were counted for either cytosolic or punctum localization of the VHL protein. The fold change in puncta number was then deduced by normalization to the *WTmultiControl* population.

5.2 Is the 100% sequence identity among the 11 copies of $tRNA^{Arg}_{UCU}$ gene family driven by selection or a neutral consequence of gene conversion?

One of the major observations in the first part of this thesis is the striking conservation of $tRNA^{Arg}_{UCU}$ copies. This naturally leads to a more global question: why are all iso-acceptor tRNA families conserved in sequence and what are the evolutionary mechanisms that govern this phenomenon? As described above, the $tRNA^{Arg}_{UCU}$ family consists of 11 copies that are 100% identical in sequence. There are two feasible hypotheses for this conservation. First, the observed sequence is maintained by selection that eliminates from the population all individuals with any potential mutations that would be deleterious. Alternatively, the sequence conservation is maintained by a concerted evolution process that utilizes gene conversion events among all 11 $tRNA^{Arg}_{UCU}$ copies. The second hypothesis implies neutrality as it suggests that 100% sequence identity is not due to selection but is rather a consequence of a neutral process.

To differentiate between these two possibilities, we created a novel methodology to assess the evolutionary processes that can occur in multi-copy gene families (see figure 10 that illustrates our workflow). In our system, we bombard the cell with a potential invading sequence (that carry mutations at various levels of predicted neutrality) on a plasmid that could replace the genome copies if they introduce neutral or near neutral mutations.

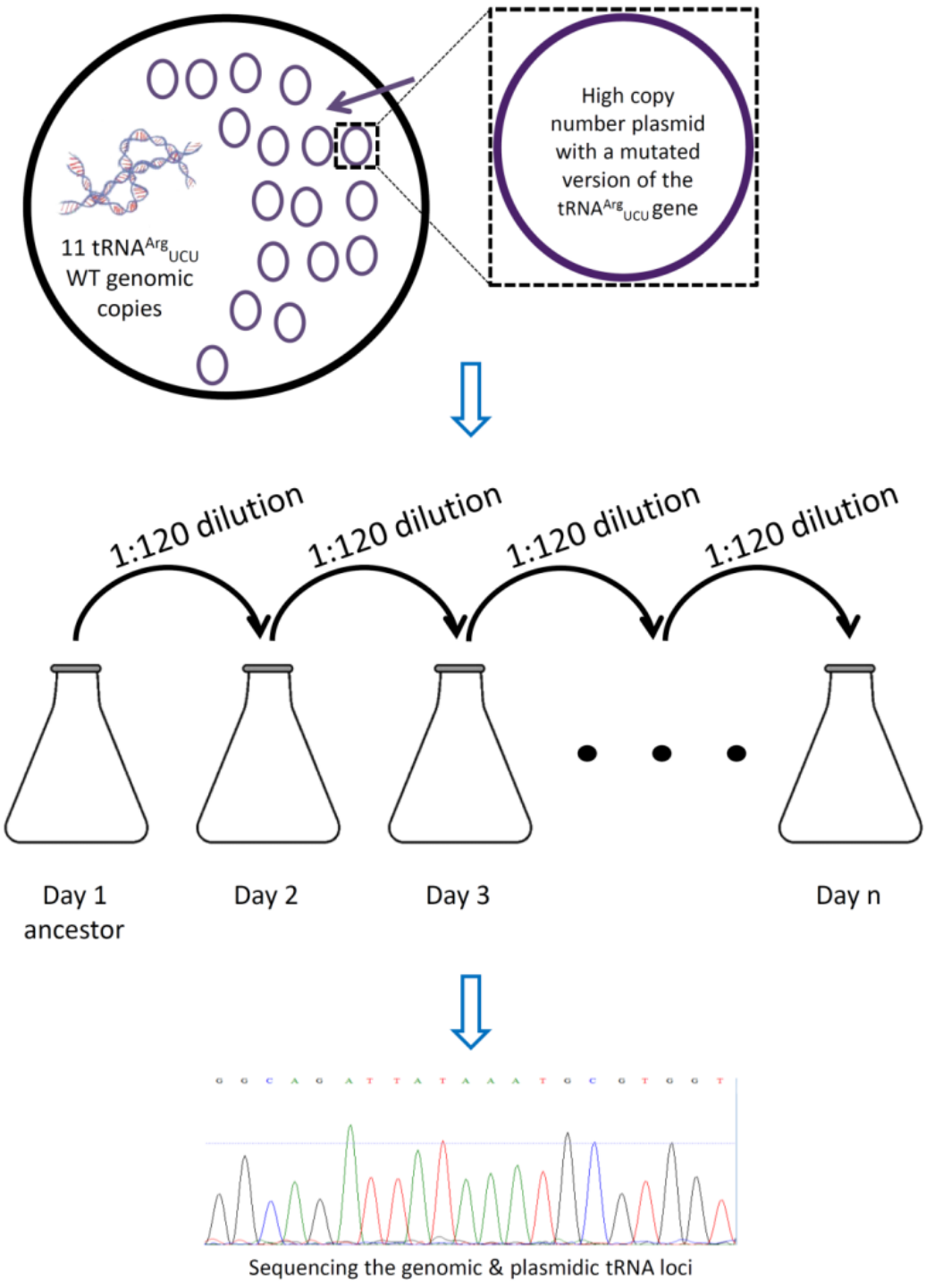


Figure 10 –Overview of our methodology

Our methodology utilizes mutagenesis process that is followed by high-throughput lab evolution experiments. Specifically, various mutated copies of tRNA$^{Arg}_{UCU}$ gene were cloned into a high copy number plasmid. Then, the plasmids were transformed into either a control strain or strains carrying mutations in recombination pathways. We allow the strains to undergo a lab-evolution in tens of parallel cultures. At the end of the evolutionary period we sequence all genomic copies of the gene as well as the dozens of mutated plasmid copies. If a neutral recombination mechanism is responsible for preserving sequence identity, it will be visualized as homogenization of the populations of the endogenous and plasmid genes. In contrast, if selection is responsible for the maintenance of the high sequence identity, then the genomic sequence will be preserved.

Cloning $tRNA^{Arg}_{UCU}$ into a high copy number plasmid

The first step towards the establishment of our approach was to clone the $tRNA^{Arg}_{UCU}$ gene to a high copy number plasmid. Since high numbers of cloning procedures had to be carried during the establishment of our systems, I imported a previously published cloning method to the lab, which was not used here before. This novel cloning procedure, termed "Restriction-Free (RF) Cloning", is based on PCR amplification of a DNA fragment that is subsequently used as a "mega-primer" for the linear amplification of the vector. This procedure results in the insertion of the original DNA fragment into the plasmid[67]. Since RF cloning is performed by two sequential PCR procedures only, it can be used in a high-throughput manner and many cloning reactions can be accomplished simultaneously. Thus, the RF cloning methodology was suitable for executing ~35 cloning procedures to create all the plasmids in this experiment (see below).

To this end, we cloned $tRNA^{Arg}_{UCU}$ into the high copy number plasmid pRS426 with a URA3 resistance marker[68]. Gene conversion is hypothesized to be involved in the evolution of this gene and its rates are known to be influenced by the length of the homology. For this reason we created 4 different plasmids with ranging number of up- and down-stream sequences of $tRNA^{Arg}_{UCU}$. First, only the 72bps, identical in all the 11 copies of $tRNA^{Arg}_{UCU}$, were cloned into pRS426. Second, we cloned these 72bps with addition of 25 up- and down-stream bps taken from the $tRNA^{Arg}_{UCU}$ copy on chromosome *XI*, tR(UCU)K. The third version of the plasmid is the same as the second, only with 200 up- and down-stream bps taken from the copy tR(UCU)E. The fourth plasmid also contains $tRNA^{Arg}_{UCU}$ with 200 up- and down-stream bps, only from the tR(UCU)K copy. We used these two copies of $tRNA^{Arg}_{UCU}$ since their deletion results in different growth phenotype, as was shown in the tRNA deletion project that is carried by Dr. Zohar Bloom at the lab. We termed these plasmids p1-pRS426-tR(UCU), p2-pRS426-tR(UCU)K-25, p3-pRS426-tR(UCU)E-200 and p4-pRS426-tR(UCU)K-200, respectively.

Since plasmids may be lost during the growth of cells[82–84], we decided to add a YFP gene to each of the plasmid in order to follow the process. Assuming that the YFP production correlates with the actual plasmid copy number, this should have

allowed us to follow the plasmid copy number in our strain throughout the lab-evolution experiment. Thus, to each of the above four plasmids, we cloned a YFP gene driven by the endogenous promoter of the ribosomal protein RPL3 using, again, the RF cloning methodology. We termed each of the new plasmids the following: p1-pRS426-tR(UCU)-YFP, p2-pRS426-tR(UCU)K-25-YFP, p3-pRS426-tR(UCU)E-200-YFP and p4-pRS426-tR(UCU)K-200-YFP.

Mutating the plasmidic-based $tRNA^{Arg}_{UCU}$

The next step toward building of our experimental system was to decide which mutation to introduce to the sequence of $tRNA^{Arg}_{UCU}$. Obviously, any mutation that should be subjected to a strong purifying selection is not suitable to our needs. For example, a mutation that changes the anti-codon triplet from arginine to any other amino-acid results in changes in the genetic code and mis-incorporation of arginine instead of a different amino-acid. Thus, mutations that severely hamper the function of the tRNA are expected to get eliminated from the population merely by selection. Instead we were looking for two types of mutations: neutral mutation with at most low fitness effect, or mutations that are deliberately predicted to exert an effect that is nonetheless not detrimental. To deduce possible relevant mutations within the $tRNA^{Arg}_{UCU}$ gene that would belong to either of these two types, we used the following 5 criterions.

Criterion 1 – structural proximity to arginyl-tRNA synthetase

All tRNA molecules have to be loaded with the correct amino-acid in order to function properly. This is achieved in the cell with the 20 amino-acyl tRNA synthetase genes that are the connecting point between the nucleotide and amino-acid languages[85]. Indeed, tRNA synthetase genes were uniquely shaped to recognize an amino-acid and its adequate tRNA molecule. Hence, any mutation we wished to introduce to $tRNA^{Arg}_{UCU}$ must not damage the ability of arginyl-tRNA synthetase to load an arginine amino acid on the tRNA. To validate the important contact sites between the tRNA and the arginyl-tRNA synthetase we used published structural data of *S. cerevisiae* arginyl-tRNA synthetase with $tRNA^{Arg}_{ACG}$[86]. Unfortunately, there

is no report for the structure of $tRNA^{Arg}_{UCU}$ so we used the only available structure published in the Protein Data Bank (PDB) for a yeast arginine tRNA with an arginyl synthetase, namely $tRNA^{Arg}_{ACG}$. Figure 11 demonstrates the distance of each $tRNA^{Arg}_{ACG}$ nucleotide to its nearest arginyl-tRNA synthetase amino acid. A limitation of this criterion is the assumption that all relevant positions for the recognition process of the tRNA by the synthetase are equal between all arginine tRNA families.

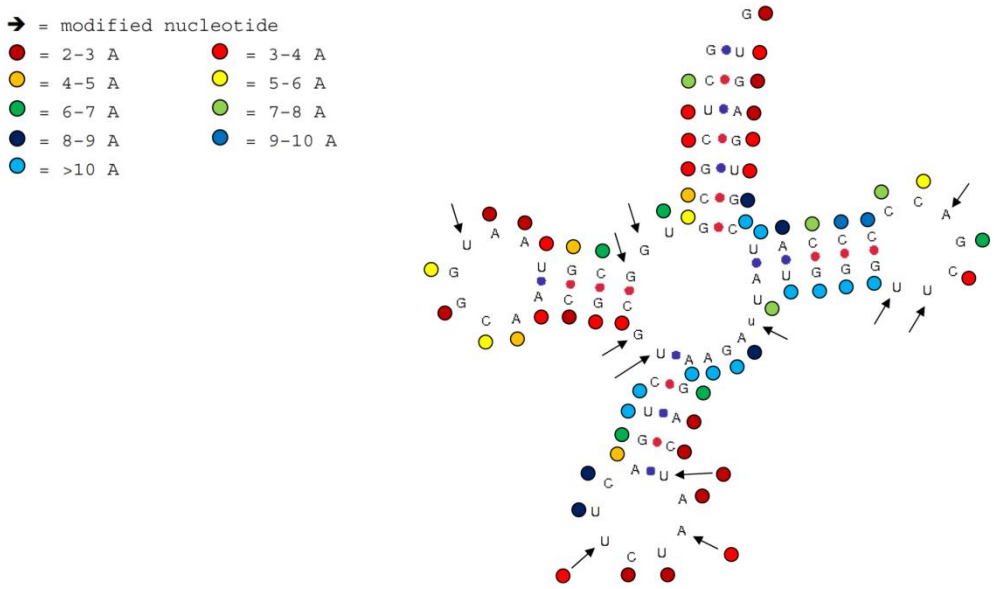


Figure 11 – distances of each position between $tRNA^{Arg}_{ACG}$ and arginyl-tRNA synthetase of S. cerevisiae
A colored cycle, indicative to the distance to the arginyl-tRNA synthetase is given near each nucleotide. Note that although the sequence is written for $tRNA^{Arg}_{UCU}$, the distances were calculated for $tRNA^{Arg}_{ACG}$ for which data was available.

Criterion 2 – *in vitro* efficiency of Arginine loading on the tRNA molecule

Complementary to the first criterion, we took into account mutations that were shown to impair the *in vitro* loading rate of arginine to $tRNA^{Arg}_{UCU}$. Figure 12 is a published[87] work in which single mutations variants to the tRNA were incubated with arginine and arginyl-tRNA synthetase. Amino-acylation was measured for each variant. C35 was found to be the most important position for the amino-acylation rate, mutations in other positions gave moderate reduction, while other mutates did not change its *in vitro* rates at all.

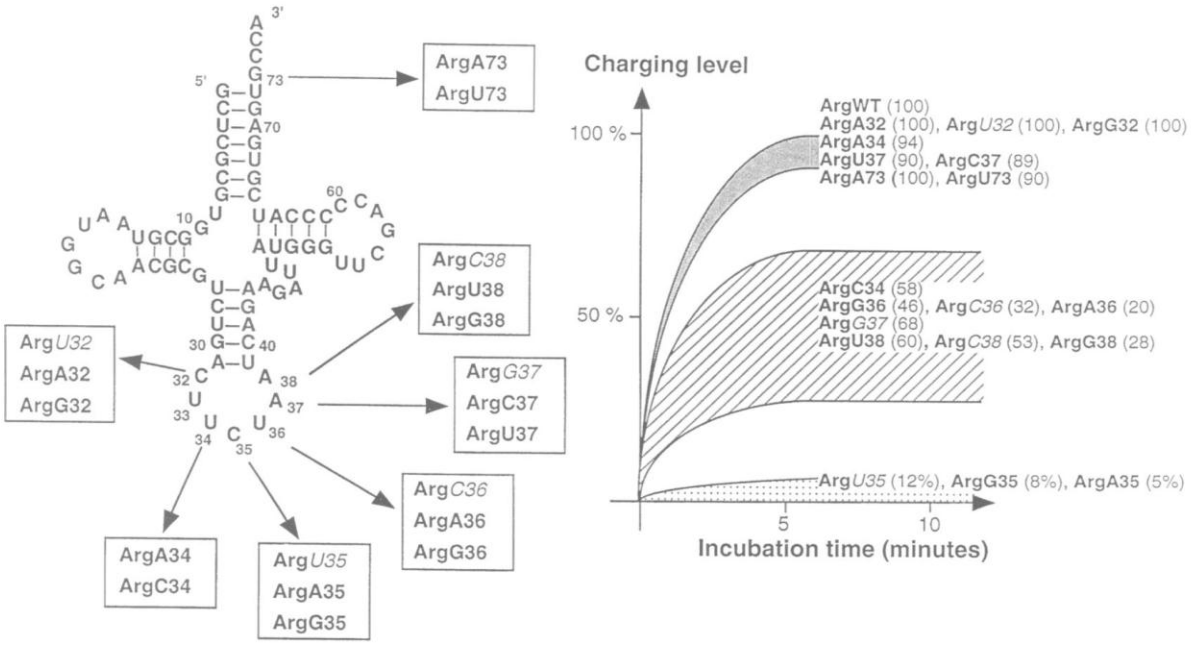


Figure 12 – Sequences and charging levels of $tRNA^{Arg}_{UCU}$ variants
(a) Nomenclature of the tRNA derived variants. **(b)** Schematic summary of the arginylation curves observed for the transcripts. Values in brackets correspond to the final charging levels.

Criterion 3 - conservation among other arginine tRNAs

Since all tRNAs have the same typical structure[36,37], we assume that the conservation among all the arginine tRNAs in *S. cerevisiae* may mark the important positions for the proper function of the tRNA. Figure 13 shows the multiple sequence alignment of all 4 arginine tRNA families[88]. The first 3 families (harboring the anti-codons CCU, ACG & UCU), which are responsible for the decoding of ~95% of the arginine codons, show 53% identity. The last arginine tRNA with the anti-codon CCG is distinguished from the other three and shares only 33% of common nucleotides.

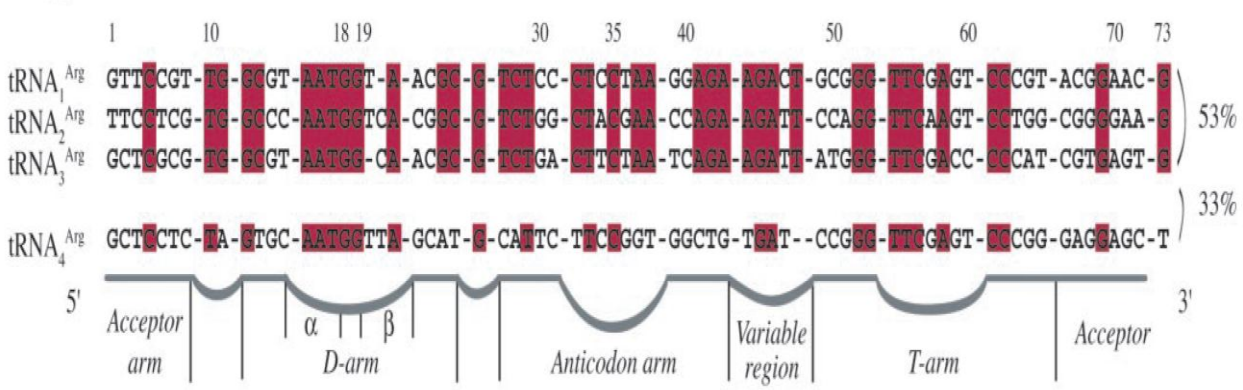


Figure 13 – MSA of the four arginine tRNA families in *S. cerevisiae*
The tRNA sequences of all 4 families (anti-codons CCU, UCU, ACG & CCG) were aligned. The secondary structural domains are indicated below the primary sequences. It can be seen that the CCG family is more distant the other three families and is actually more resembles to an aspartate tRNA family[88].

Criterion 4 – conservation among $tRNA^{Arg}_{UCU}$ genes of 10 yeast species

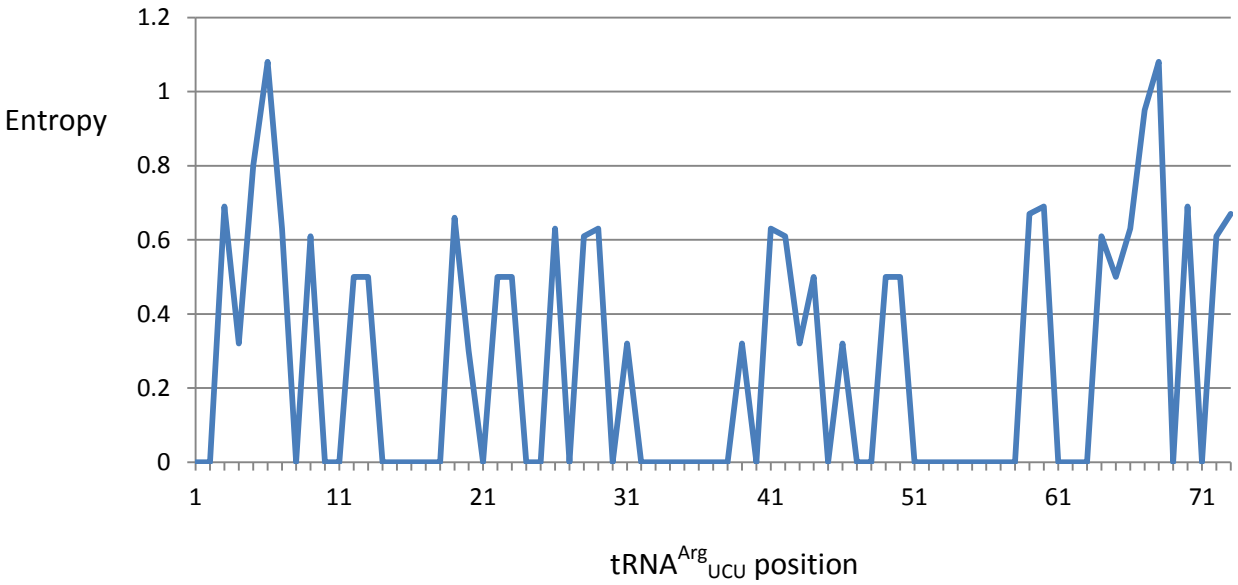
In addition to the conserved positions of the 4 arginine tRNA families in *S. cerevisiae*, we hypothesized that conservation of $tRNA^{Arg}_{UCU}$ specifically among different yeast species can also reveal functional positions. Figure 14 shows the multiple sequence alignment of 10 different sequences taken from yeast species of $tRNA^{Arg}_{UCU}$. Out of the 72bps, about 50% demonstrate complete conservation.

A.

```
cerevisiae              GCTCGCGTGGCGTAATGGC-AACGCGTCTGACTTCTAATCAGAAGATTATGGGTTCGACCCCCATCGTGAGTG
Candida_glabrata        GCTCGCGTGGCGTAATGGC-AACGCGTCTGACTTCTAATCAGAAGATTATGGGTTCGACCCCCATCGTGAGTG
Eremothecium_gossypii   GCTCGCGTGGCGTAATGGC-AACGCGTCTGACTTCTAATCAGAAGATTGTGGGTTCGACCCCCACCGTGAGTG
Kluyveromyces_lactis    GCTCGCGTGGCGTAATGGC-AACGCGTCTGACTTCTAATCAGAAGATTGTGGGTTCGACCCCCACCGTGAGTG
Debaryomyces_hansenii   GCCTGCGTAGCGTAATGGTTAACGCGTTTGACTTCTAATCAAAAGATTGCGGGTTCGACTCCCGCCGTGGGTT
Yarrowia_lipolytica     GCCCGAGTAGCGCAATGGTTAACGCGTTGGACTTCTAATCCAAAGATTGTGGGTTCGAGTCCCACCCCGGGTG
Aspergillus_fumigatus   GCCCTGCTGGCGCAATGG-TAGCGCGTCAGACTTCTAATCTGAAGGTTGTGGGTTCGACCCCCACGTAGGGCT
pombe                   GCTCCCGTGGCCTAATGGCTAGGGCATTTGACTTCTAATCAAGGGATTGTGGGTTCGAGTCCCGCCGGGAGCT
Cryptococc_neoformans   GCCCGTGTGGCCTAATGGTTAAGGCTTCTGACTTCTAATCAGAGGATTGTGGGTTCGAGTCCCACCTTGGGCT
Encephalitozo_cuniculi  GCCCTGATAGCGTAATGGCTAACGCGTCTGCCTTCTAAGCAGAAGACTGCGGGTTCGAGTCCCGCTCAGGGTG

Legend:
Red - completely conserved
Purple - partially conserved
Black - least conserved
Highlight - anti-codon (positions 34-36)
```

B.



C.

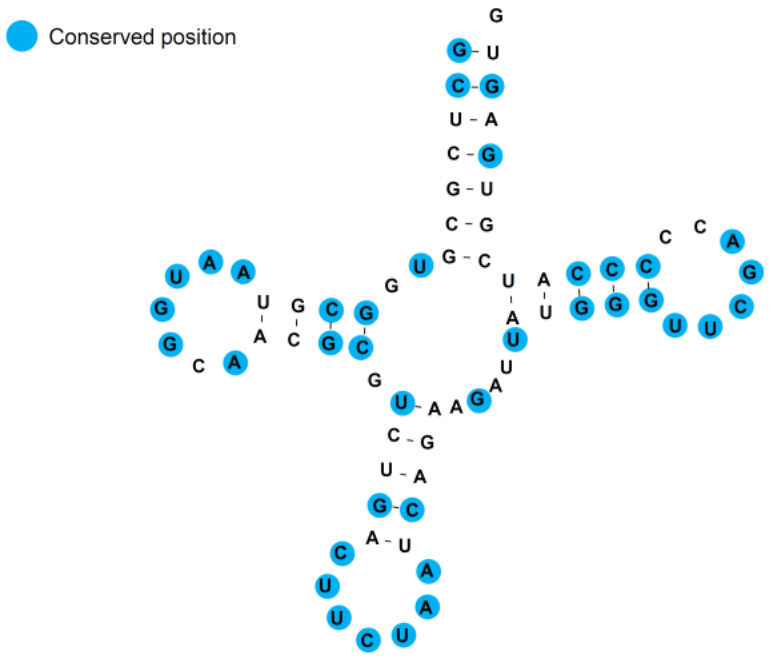


Figure 14 – MSA of 10 tRNA$^{Arg}_{UCU}$ genes from 10 different yeast species
(a) The MSA, performed with MATLAB. **(b)** Entropy scores per position in the alignment. The entropy was calculated with the following formula: $\sum_{n=1}^{4} p(n) * \log(p(n))$ where n represents the four possible nucleotides and p is the probability to see a given nucleotide at a specific position. **(c)** The sequence of *S. cerevisiae* tRNA$^{Arg}_{UCU}$ with all completely conserved nucleotides among yeast species marked with blue circles.

Criterion 5 – conservation among all *S. cerevisiae* tRNA genes

Naturally, any position that is conserved throughout all the tRNAs is important to the function of these genes as a mediators of amino-acids. Figure 15, which was recently published[89], describes the conservation among all tRNA sequences of *S. cerevisiae*. Strikingly, the only two domains found to be conserved throughout all genes are the A and B boxes – the transcription control motifs that mediate RNA polymerase III transcription and reside within the transcribed region itself[36,37]. Any mutation in these conserved sites is expected to hamper the transcription of the tRNA gene. Thus, we excluded any conserved position in these boxes from being mutated in our experiment.

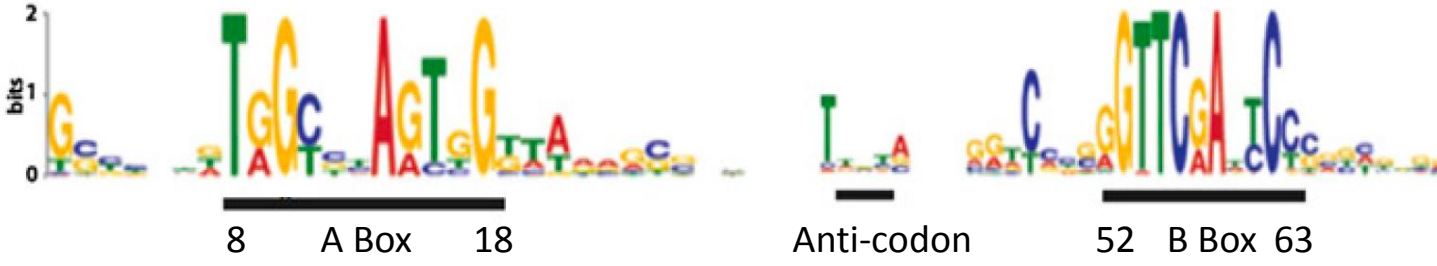


Figure 15 – sequence logo of all *S. cerevisiae* tRNA genes
Sequences of all of the tRNA genes from *S. cerevisiae* were aligned and then subjected to Sequence Logo. Position 1 reflects the first nucleotide of the mature tRNA molecule. The A and B boxes and the anti-codon are indicated. The A and B boxes are the most conserved domains among the tRNA genes.

Overlapping positions of all criteria

Each criterion yielded a group of positions along the $tRNA^{Arg}_{UCU}$ sequence that are suggested to be less functional and thus might not be preserved by purifying selection. We then looked for the positions in which a mutation is likely to be neutral. These positions are expected to be the intersection of all criterions. Figure X is a calculated Venn diagram showing the distribution of positions among the different criterion. The joint positions by all criteria are: 7, 19, 49, 50, 59, 60, 64, 65, 66 & 67. We excluded criterion 2 from the Venn diagram since data is available only for few positions.

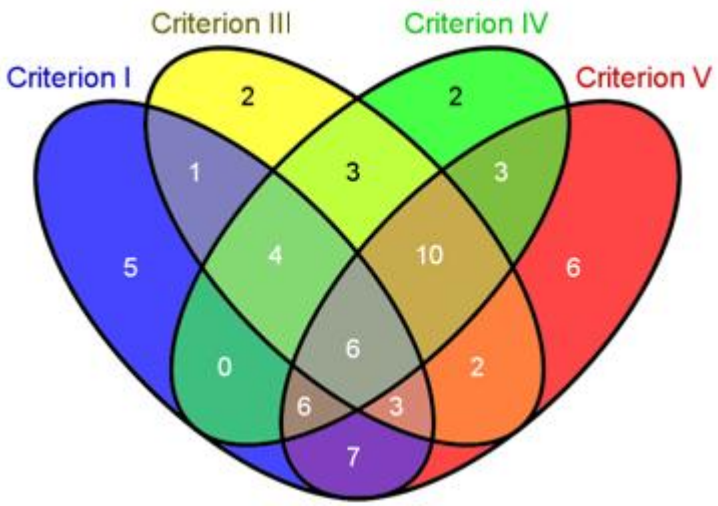


Figure X – Venn diagram for the intersection of the mutual position for 4 Criterions
A Venn diagram for criterion I, II, IV & V is demonstrating the shared positions along the $tRNA^{Arg}_{UCU}$ sequence among the criterions. Criterion III was left out as data is available for few positions only.

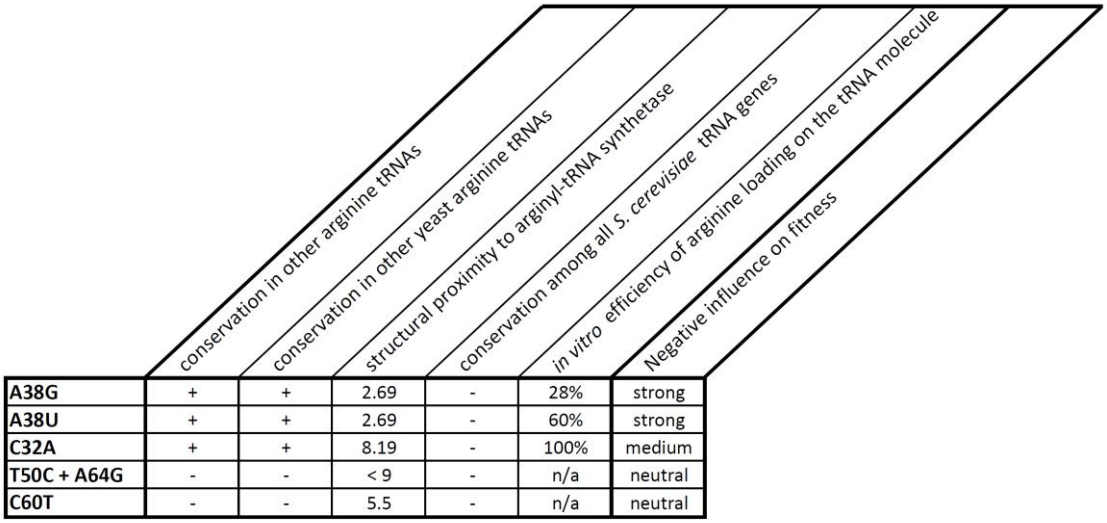
Introducing 5 mutation to $tRNA^{Arg}_{UCU}$

Considering all of the above criteria, we decide to choose the following five mutations that each represents either a predicted neutral or non-neutral substitution (Figure 16).

- **Point mutation #1** - A38G - This mutation is located in the anti-codon loop, in the vicinity of the anti-codon itself (positions 34-36). The mutation was shown *in vitro* to decrease the rate of arginine loading to 28%. This position is also conserved in two other $tRNA^{Arg}$ families. In addition, this position is completely conserved in 10 sequences from 10 different yeast species. Lastly, the distance of this position in $tRNA^{Arg}_{ACG}$ from the arginyl-tRNA synthetase is 2.69Å. Thus, this mutation is not considered to be neutral and is hypothesized to be selected against.
- **Point mutation #2** - A38U- This mutation is at the same location as the previous one. The only difference is that it was shown to decreases the rate of arginine loading to 60% only. Thus, any purifying selection that may act on this mutation should be less strong than mutation #1.
- **Point mutation #3** - C32A – This mutation is located in the anti-codon loop. On the one hand, it was shown *in vitro* that this mutation does not affect the rate of arginine loading. On the other hand, this position is conserved in two

other $tRNA^{Arg}$ families. In addition, this position is completely conserved in 10 sequences from 10 different yeast species. Lastly, the distance of this position in $tRNA^{Arg}_{ACG}$ from the arginyl-tRNA synthetase is large: 8.19Å. We chose to avoid replacing C with U both in order to prevent it from being complementary to position A38 and to prevent a UUU sequence in the tRNA- what might have changed the secondary structure severely.

- **Point mutation #4** - T50C + A64G – These mutations are hypothesized to be neutral. They are located on the T-arm of the tRNA and do not interact with the synthetase (distances larger than 9A). These positions are only partially conserved in the other $tRNA^{Arg}$ families (there is Purine-Pyrimidine conservation). In addition, these positions are only partially conserved in 10 sequences from 10 different yeast species. Since these two positions complement each other in the secondary structure model, we decided to maintain the complementation. There were two possibilities for mutations in these positions. First, switching the order of the nucleotides. This way, the strength of the chemical bonds should be the same and hopefully the structure is minimally influenced in terms of flexibility and strength. Yet, this option violates the Purine-Pyrimidine conservation in these two positions. Second, and the chosen option, is to mutate the positions according to other $tRNA^{Arg}$ genes and maintain the Purine-Pyrimidine conservation.
- **Point mutation #5** - C60T – This mutation is located in the T loop of the tRNA and should be neutral. These position is only partially conserved in the other $tRNA^{Arg}$ families (there is Purine-Pyrimidine conservation) and is partially conserved among the other yeast species (there is Purine-Pyrimidine conservation). We chose to replace C with T since this leaves the Purine-Pyrimidine conservation intact and still maintain the non-complementary state of this position with position U54.



	conservation in other arginine tRNAs	conservation in other yeast arginine tRNAs	structural proximity to arginyl-tRNA synthetase	conservation among all *S. cerevisiae* tRNA genes	*in vitro* efficiency of arginine loading on the tRNA molecule	Negative influence on fitness
A38G	+	+	2.69	-	28%	strong
A38U	+	+	2.69	-	60%	strong
C32A	+	+	8.19	-	100%	medium
T50C + A64G	-	-	< 9	-	n/a	neutral
C60T	-	-	5.5	-	n/a	neutral

Figure 16 – Summary of the mutation introduced to the plasmidic tRNA$^{Arg}_{UCU}$
This table summarizes the location and type of all 5 point mutations in our experiment according to all 5 criterions.

Each of the above mutations was introduced separately to each of the plasmid containing the tRNA$^{Arg}_{UCU}$ gene using the RF cloning method. Thus, additional 20 plasmids were created and given the following names:

1 – p1-pRS426-tR(UCU)-YFP -mut1

2 - p1-pRS426-tR(UCU)-YFP -mut2

3 - p1-pRS426-tR(UCU)-YFP -mut3

4 - p1-pRS426-tR(UCU)-YFP -mut4

5 – p1-pRS426-tR(UCU)-YFP -mut5

6 – p2-pRS426-tR(UCU)K-25-YFP -mut1

7 – p2-pRS426-tR(UCU)K-25-YFP -mut2

8 – p2-pRS426-tR(UCU)K-25-YFP -mut3

9 – p2-pRS426-tR(UCU)K-25-YFP -mut4

10 – p2-pRS426-tR(UCU)K-25-YFP -mut5

11 - p3-pRS426-tR(UCU)E-200-YFP -mut1

12 - p3-pRS426-tR(UCU)E-200-YFP -mut2

13 – p3-pRS426-tR(UCU)E-200-YFP -mut3

14 - p3-pRS426-tR(UCU)E-200-YFP -mut4

15 - p3-pRS426-tR(UCU)E-200-YFP -mut5

16 - p4-pRS426-tR(UCU)K-200-YFP-mut1

17 – p4-pRS426-tR(UCU)K-200-YFP-mut2

18 – p4-pRS426-tR(UCU)K-200-YFP-mut3

19 - p4-pRS426-tR(UCU)K-200-YFP-mut4

20 – p4-pRS426-tR(UCU)K-200-YFP-mut5

Tackling potential effect of the recombination pathways on the evolutionary dynamics in the experiment

The lab-evolution experiment was performed on a WT background of BY4741. Since the most probable homologous recombination pathway that could be involved in a putative concerted evolution of duplicated genes is gene conversions (see introduction), we decided to add two additional genetic backgrounds to our experiment, in which homologous recombination is hampered.

The *RAD52* genes group was identified by the requirement for the repair of ionizing radiation-induced DNA damage in *S. cerevisiae*. These genes are central to the process of homologous recombination and are highly conserved among eukaryotes[12]. The two most characterized and studied genes are the yeast *RAD52* and *RAD51*[11,12]. Usually, once a double strand break (DSB) has occurred *RAD52* assumes its role as a recombination mediator and forms a complex with Rad51. This allows *RAD51*'s initial formation of the presynaptic filament on single stranded (SS) DNA. Then, further polymerization of additional Rad51 molecules is carried and the process of homologous recombination is completed via one of four different pathways (see figure 17 for more details). Although *RAD52* is required for all homologous recombination pathways, there are 2 pathways that can be carried out in a *RAD51*-independent manner, single strand annealing and break induced replication[11,12]. Indeed, deletion of *RAD52* in *S. cerevisiae* results in a severe reduction in homologous recombination rates, including gene conversion and single strand annealing events[90].

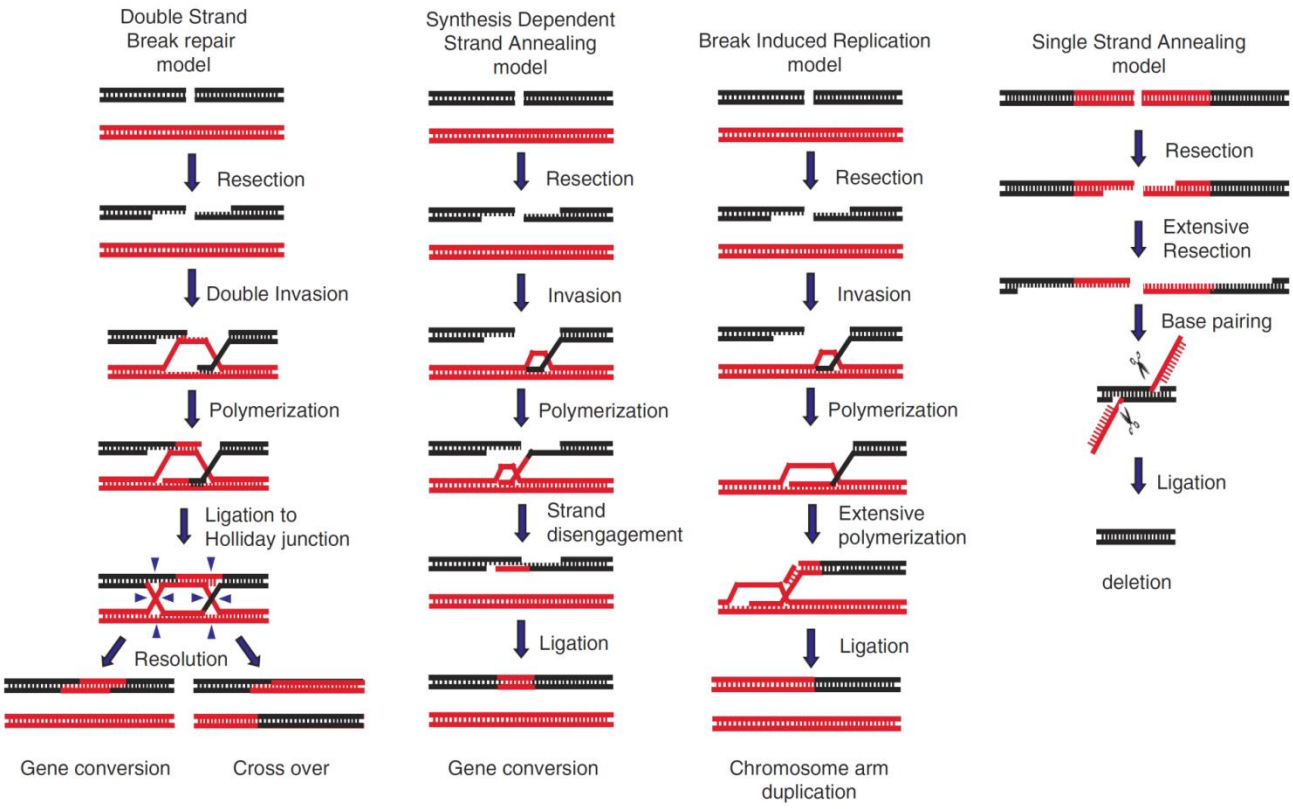


Figure 17 – Schematic representation of the different DSB repair models[13]
Homologous recombination can act in 4 pathways. In the Double Strand Break Repair model, a double Holliday junction (HJ) can be resolved and generate either crossover or gene conversion products. In the Synthesis Dependent Strand Annealing model, D-loop extension and invading strand displacement produce a gene conversion event. In the Break Induced Replication model, the invading strand continues DNA synthesis to the end of the DNA molecule, producing a duplication of the chromosome arm. In the Single Strand Annealing model, a DSB made between two direct repeats results in deletion of one of the repeats.

To investigate the lab-evolution consequences of our new plasmids on genetic backgrounds that reduce recombination rates, we decided to transform our set of plasmids to both Δ*rad51* and Δ*rad52* strains in addition to the wild-type cells. Since different genetic backgrounds may lead to different evolutionary outcomes, we wanted to minimize the primary differences among our three strains (WT BY4741, Δ*rad51* and Δ*rad52*). To this end, we transformed a G418 resistance marker that replaced either the *RAD51* or *RAD52* loci on the background of the WT strain. Figure 18 illustrates the growth curves of these two strains compared to the growth of the WT. In line with previous reports[91], we demonstrate that both these strains grow slower than the WT, with Δ*rad52* growing even slower than Δ*rad51*.

A. B.

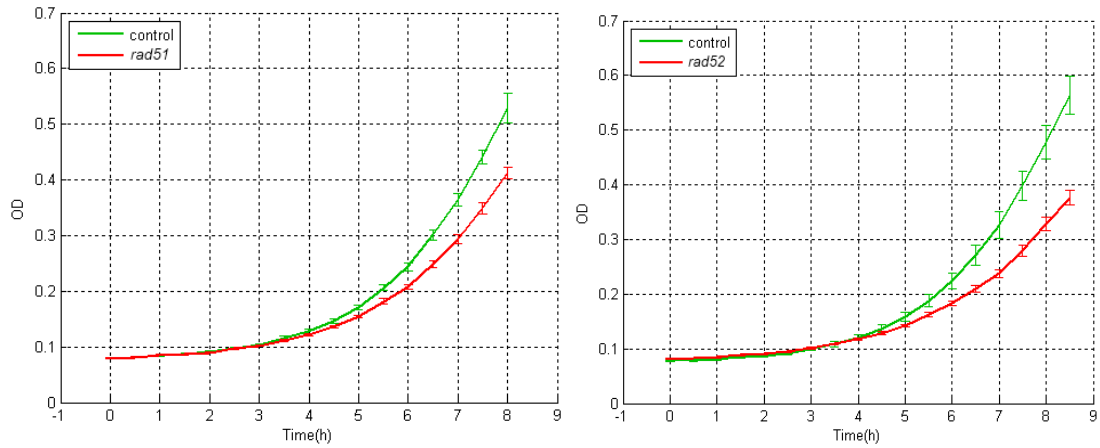
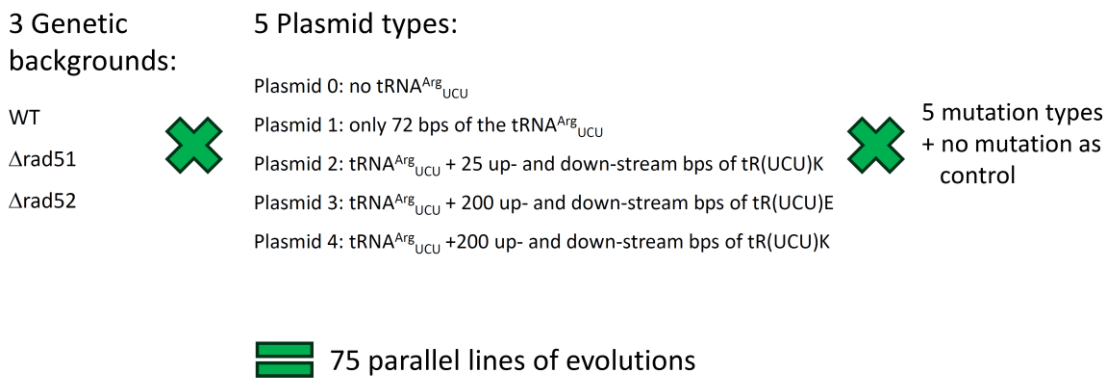


Figure 18 – growth curves of Δ*rad51* and Δ*rad51* cells compared to the WT
Both growth experiments were performed in a 96well plate ordered in a checker board manner with control and deletion strains spanning 48wells each. OD measurements were taken once every 45 minutes. **(a)** Control (green curves) vs. Δ*rad51* (red curve). **(b)** Control (green curves) vs. Δ*rad52* (red curve).

Our lab-evolution experiment encompassed a total of 75 different strains and was carried out in a 96well plate format

Once we obtained the 20 different plasmids (each containing the $tRNA^{Arg}_{UCU}$ gene with ranging up- and down- sequences and a different mutation in the sequence) together with the two additional genetic backgrounds (Δ*rad51* and Δ*rad52*), we needed to create all the combinatorial strains. Hence, I utilized a high-throughput transformation protocol that allowed for the transformation of our 75 needed strains in a single experiment. The transformation protocol was carried in a 96well plate format and resulted in the creation of the following strains:



After the 75 strains were created, we started the lab-evolution experiment, which was performed in a single 96well plate. Figure 19 is the map of all strains, indicating their positions on the lab-evolution plate. The volume of each culture was 150µl and each day, 2µl were diluted into a fresh medium. The duration of the lab-evolution experiment was ~260 generations.

	1	2	3	4	5	6	7	8	9	10	11	12
A	WT-p0	WT-p1	WT-p2	WT-p3		Δrad51-p0	Δrad51-p1	Δrad51-p2		Δrad52-p0	Δrad52-p1	Δrad52-p2
B	WT-p4	WT-p1-mut1	WT-p1-mut2			Δrad51-p3	Δrad51-p4	Δrad51-p1-mut1		Δrad52-p3	Δrad52-p4	Δrad52-p1-mut1
C	WT-p1-mut3	WT-p1-mut4	WT-p1-mut5		Δrad51-p1-mut2	Δrad51-p1-mut3	Δrad51-p1-mut4			Δrad52-p1-mut2	Δrad52-p1-mut3	Δrad52-p1-mut4
D	WT-p2-mut1	WT-p2-mut2	WT-p2-mut3			Δrad51-p1-mut5	Δrad51-p2-mut1		Δrad52-p1-mut5	Δrad52-p2-mut1	Δrad52-p2-mut2	Δrad52-p2-mut3
E	WT-p2-mut4	WT-p2-mut5	WT-p3-mut1		Δrad51-p2-mut2	Δrad51-p2-mut3	Δrad51-p2-mut4			Δrad52-p2-mut4	Δrad52-p2-mut5	Δrad52-p3-mut1
F	WT-p3-mut2	WT-p3-mut3	WT-p3-mut4		Δrad51-p2-mut5	Δrad51-p3-mut1	Δrad51-p3-mut2	Δrad51-p3-mut3		Δrad52-p3-mut2	Δrad52-p3-mut3	Δrad52-p3-mut4
G	WT-p3-mut5	WT-p4-mut1	WT-p4-mut2			Δrad51-p3-mut4	Δrad51-p3-mut5	Δrad51-p4-mut1		Δrad52-p3-mut5	Δrad52-p4-mut1	Δrad52-p4-mut2
H	WT-p4-mut3	WT-p4-mut4	WT-p4-mut5		Δrad51-p4-mut2	Δrad51-p4-mut3	Δrad51-p4-mut4	Δrad51-p4-mut5		Δrad52-p4-mut3	Δrad52-p4-mut4	Δrad52-p4-mut5

Figure 19 – Map of positions of all strains during the lab-evolution experiment
The name of each strain, out of the total 75 strains in our experiment, is written according to its position on the plate during the lab-evolution. The strains were divided according to genetic backgrounds. First five strains of each genetic background were controls harboring plasmids carrying no mutation. Black squares are empty wells with only medium to control for cross-contaminations.

Sanger sequencing of the 12 $tRNA^{Arg}_{UCU}$ loci from all evolved strains

To elucidate the evolutionary dynamics and outcome of the 12 $tRNA^{Arg}_{UCU}$ in our 75 evolved strains we needed to sequence these loci. Since our mutations are not necessarily expected to be fixed in the population and may appear at a range of frequencies, we purified DNA from a sample of the entire evolved population rather than using the genome of a single clone out of the population. Using specific primers for each locus, we amplified the 12 $tRNA^{Arg}_{UCU}$ loci via PCR. In addition, to avoid any PCR biases, which may artificially alter the frequencies of SNPs in a population, the PCR amplifying the 12 $tRNA^{Arg}_{UCU}$ loci had only 24 rounds of DNA amplifications. Figure 20 shows the results of a Real-Time PCR experiment following the double-

stranded DNA concentration for all primer pairs we used. It can be seen that all reactions are still at the exponential phase of amplification after 24 rounds.

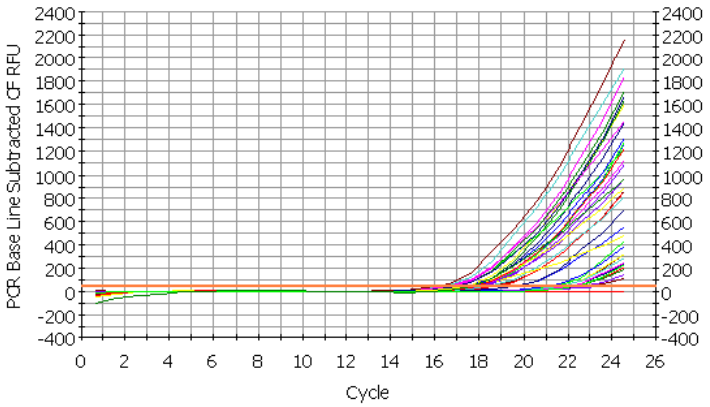


Figure 20 – RT-PCR results for all primer pairs used to amplify the 12 $tRNA^{Arg}_{UCU}$ loci of the evolved population
Each primer pair had 3 reactions for which the ds-DNA concentration was followed. It can be seen that after 24 amplification cycles, all reactions were still in the exponential phase of the amplifications. This minimizes biases caused due to low efficiency of the PCR. Non amplified curves represent negative controls.

We sent the amplified DNA fragments of all 12 $tRNA^{Arg}_{UCU}$, which were linearly amplified in the PCR, to sequencing using the Sanger sequence method that allows for discovery of a SNP at a minimal frequency of 5-10% (depending on the noise of the sequencing reaction)[92]. For lack of time, we sequenced only the 25 strains with a WT background. Since each strain was sequenced in 12 loci, we sequenced a total of 300 loci. We then went over all the sequencing results to determine whether mutations from the plasmid invaded each of the genomic loci or if the plasmidic copy returned to the WT sequence of $tRNA^{Arg}_{UCU}$. Furthermore, to avoid loss of data in cases of small SNP frequencies, we scanned by eye the chromatograms of all the 300 sequencing procedures in search for even very rare minor variants frequencies in the population. By comparing the sequencing chromatogram's highest point of the expected mutation on the expected position to the average noise level of 6 up- and down-stream positions, we could determine if at least 5-10% of the population possess a SNP in a genome copy that could originate from the plasmid, or conversely a SNP in the plasmid. No observations of a mutation were found for the genomic and plasmidic loci in all 300 reactions tested. The fact that no mutations infiltrated from the plasmid to the genome could be taken to indicate that the genomic sequence is

functional as is and that even the most neutral mutations were selected against. However, it is also possible that the efficiency of the gene conversion process in our setup was too low for fixation of mutations in the current time frame of the lab-evolution experiment (see Discussion below).

6 | Discussion

tRNAs are thought to be among the oldest biological entities, which were present in the last universal common ancestor (LUCA), and are enabling the genetic code by linking anticodon to amino acid[93]. The redundancy of the genetic code, in which most of the amino acids can be translated by more than one codon, enables the coding of the same amino acid sequence by numerous possibilities. The various codons that correspond to the same amino acid are termed synonymous codons and their corresponding tRNAs may differ in their genomic copy number. As tRNA copy number correlate to its abundance in the cell, the speed in which tRNAs are recognized by the ribosome varies[94]. Hence, this genetic structure allows evolution to fine-tune both the supply (tRNA genes) and demand (the frequency of use per codon) in translation.

In general, many tRNAs are able to pair with more than one codon, and many codons may be translated by more than one tRNA type due to wobble interactions between the codon and anti-codon. It was previously suggested that if high tRNA concentrations are required, duplicate tRNA genes are selected in order to provide a higher rate of tRNA transcription. Thus, codon frequencies and tRNA copy numbers co-evolve toward a supply vs. demand balance[95,96].

Indeed, translational selection acts to increase the frequency of preferred codons in high-expression genes. Additionally, highly used codons tend to correspond to the tRNAs that have the highest concentrations in cells[94]. However, the evolutionary scenario that leads to a change in the tRNA pool and the evolutionary path cells undergo when encountering a lasting break in the supply vs. demand balance are elusive. Moreover, whether selection acts only towards increasing the copy number of a tRNA that corresponds to an abundant codon, or can it also maintain tRNAs at relatively low copy numbers is an open question.

To reveal the importance of the balance between supply and demand, we introduced a severe genetic perturbation in a yeast strain in which the balance was broken by a singleton $tRNA^{Arg}_{CCU}$ gene deletion. Supportively, such deletion confers a growth defect to the cell compared to the WT. Lab-evolution experiments performed on this strain demonstrate that the balance was rapidly restored by mutations in

other $tRNA^{Arg}_{UCU}$ genes that, despite 21 sequence differences, compensates for the $tRNA^{Arg}_{CCU}$ deletion.

In addition, over expression of $tRNA^{Arg}_{CCU}$ challenges the protein quality control machinery and generates a folding stress in the cell. This phenomenon may result from at least three reasons. First, the over expression of $tRNA^{Arg}_{CCU}$ may mis-incorporate arginine into non-arginine codons and by that lead to proteotoxic stress. Second, other aminoacyl tRNA synthetases may load wrong amino acid to the highly expressed tRNA. Mis-loading will result in the incorporation of a different amino acid where arginine is needed. Both of these reasons are relevant to both over expression of $tRNA^{Arg}_{CCU}$ and $tRNA^{Arg}_{UCU}$. However, cells with a highly expressed $tRNA^{Arg}_{CCU}$ suffer from a greater growth defect and a more severe proteotoxic stress than their equivalent $tRNA^{Arg}_{UCU}$ cells. The third, and maybe most intriguing hypothesis, suggests that the phenotypes observed for over expression of $tRNA^{Arg}_{CCU}$ are due to the increase in translation speed of the AGG codon. Since slow translation of low-frequency codons was shown on a single gene basis to influence correct folding of proteins[62,64,80], it is possible that keeping certain tRNAs at low concentrations may be essential for proper folding of specific domains during the synthesis of a new protein. This hypothesis also explains why over expressing $tRNA^{Arg}_{UCU}$ is less harmful to the cell, as the translation speed of this tRNA is already high. Our work may have thus contributed in the first time to a genome-wide demonstration of the effect of slow codons on co-translation folding.

Our lab-evolution experiments suggest that the translational balance is restored in the cell by mutations in the anti-codon, which lead to changes in the tRNA pool. Such supply adaptations are selected in order to meet the codon demand and restore the balance. This adaptation path may explain how cells may adapt to new environmental conditions. Once a new transcription program is carried out by the cell to cope with a new stress, the codon demand changes and different codons are in greater use[96]. We propose that one of the ways cells may adapt to the disruption of translational balance in the above evolutionary scenario is by mutation in tRNA genes that alter the anti-codons. Additionally, the over expression experiments suggest that not only does selection raise the tRNA copy number corresponding to

abundant codons, it may also actively maintain the low tRNA copy number for rare codons. Selection is suggested to occur via the protein folding problems that are caused once a rare tRNA in over expressed.

The results of the first part of my thesis revealed the striking conservation among the 11 copies of $tRNA^{Arg}_{UCU}$. In addition with the observation that all tRNA families in *S. cerevisiae* have high similarity, these arguments suggest that tRNA genes may have evolved dependently in a concerted manner. The 100% sequence identity among the 11 copies could be the result of either selective pressure to maintain a desired sequence, or an active maintenance mechanism, such as recurrent recombination, which is less sensitive to the sequence fitness. To address this question, we created a novel methodology to assess the evolutionary processes that occur to multi-copy gene families. Our system utilizes a mutated copy of a tRNA gene, cloned into high copy number plasmids and transformed into either control strains or strains carrying mutations in recombination pathways. Then, we allowed the strains to undergo a lab evolution process in tens of parallel cultures. At the end of the evolutionary period we sequenced all genomic copies of the gene as well as the dozens of mutated plasmid copies. Our expectations were that if a recombination mechanism is responsible for preserving sequence identity, it will be manifested by infiltration of plasmid-born sequence variants into the genomic copies of tRNA genes. In contrast, if selection is responsible for the maintenance of the high sequence identity, then the genomic sequence is preserved.

As explained in the results section, we did not observe any nucleotide change in either the genomic tRNA copies or the plasmidic ones. Seemingly, this results support the hypothesis that the sequence identity of the tRNA copies is maintained by selection. However, our ability to make such conclusions is restricted due to several limitations. First, if indeed selection maintains a specific sequence for all tRNA copies, it is interesting to note that the plasmidic tRNAs did not revert back to the WT sequence. Since the plasmidic copies contribute mutate tRNA molecules to the specific pool of the $tRNA^{Arg}_{UCU}$, these molecule may influence the fidelity of translation. The observation that the plasmidic copies were not changed back to the

WT sequence may stem from the fact that the existence of the 11 WT tRNA copies is sufficient to generate a well-functioning translation process.

Second, our lab-evolution experiment lasted for ~260 generations only, a period of time which might not be a sufficiently long for the gene conversation process to act on the genomic tRNA copies. The rate of gene conversion events for homology of ~70bps was reported to be 10^{-9} events per cell per generation[97]. Since there are ~10^{7} divisions per day in each tube in our lab evolution, the expected number of days for a single gene conversion event is 100 days. However, under the assumption of ~20 plasmidic copies and 11 genomic genes, the number of options for a gene conversion event between a genomic locus to a plasmidic one is 20*11 = 220. Thus, the rate of gene conversion is adjusted to ~10^{-7} events per cell per generation, which leads to an expected value of 1 day for a gene conversion event in each test tube. Additionally, since the cultures were diluted each day in a factor of 75, this means that the chance of a single cell, e.g. a cell that underwent a gene conversion even, to be sampled into the fresh medium of the following day is 1/75 (assuming neutrality of the event) with an expected value of 75 days. Thus, 75 days are needed in order for a gene conversion event between a genomic tRNA copy to a plasmidic one to happen and to be sampled during the dilution of the culture in the lab-evolution experiment. Since the lab-experiment lasted only 40 days, it is not within the expected time phrase of the experiment that a change should have been seen. Moreover, to the above 75-day time period, the time for fixation needs to be added. Under the assumption of neutrality, any event that leads to the invasion of a mutation to a genomic copy would take 4*Ne generation to fixate in the population. Our estimated effective population size (NE) is ~$2*10^{5}$ cells, which obviously means that not enough time had passed for any such gene conversion event to fixate.

Third, in the case that gene conversion rates are higher among tRNA genes, it is plausible that invasion of the mutations from the plasmidic copies to the genome occurred. Yet, we could not detect these events since the frequency of their decedents did not rise to our detection threshold of about 5-10% by Sanger sequencing. Finally, it is possible that gene conversion acts differently between genomic copies to plasmidic ones in comparison to gene conversions among the

genomic copies themselves. This possibility implies that the mutations on the plasmidic copies could not infiltrate the genomic loci even in the case that the sequence identity among the tRNA copies is maintained by gene conversion.

In conclusion, the aim of my thesis was to reveal the mechanisms driving the evolution of tRNA multi-copy gene families. In the first part, we demonstrated the flexibility of the tRNA pool and the utilization of mutations in the anti-codon in order to restore translational balance in the cell. In the second part of the thesis, we have generated a novel system to understand whether tRNA copies of the same family evolve dependently, hence influenced by an evolutionary mechanism that can create high sequence identity among family members. The future application of our system may teach us about the interplay between selection and drift- the two main acting forces in evolution. If sequence identity due to gene conversion is a wide phenomenon, then it could constitute a novel example for conservation that is not due to selection, but is rather due to neutral processes.

7 | Future plans

There are some future aims that we plan to pursue in order to further extend our understanding of the evolution of the tRNA pool. First, in order to show the generality of the purifying selection on rare tRNA genes, we plan to transform yeast cells with an over expression plasmid harboring other rare tRNAs. We expect to observe both growth defects and increased proteotoxic levels in these strains compared to controls in which we over-express abundant tRNAs coding for the same amino acid. Additionally, we plan to deepen our understanding regarding the mechanisms that confer the phenotypes for the over expressed tRNA strains. We will design an experiment for each of the hypotheses (mis-incorporating of arginine, mis-loading of other amino acids on top of the over expressed tRNA and increasing the translation speed of rare tRNAs) in order to learn about the contribution of each of these options to the observed phenotypes. Moreover, to demonstrate that tRNA may indeed evolve by mutations in their anti-codons, we will scan all annotated tRNAs in various organisms for cases in which tRNAs show increased similarity to tRNAs with a different anti-codon rather than for other members of their family. This systematic search may reveal scenarios in the evolution of the tRNA pool that resulted in an anti-codon swap of a tRNA gene.

Separately, we plan to continue the efforts and investigate whether tRNA multi-copy families evolve dependently. Since the lab-evolution experiment did not result in changes in the examined loci, we plan to continue evolving the strains further. Providing more evolution time may reveal phenotypes that are now hidden. In addition, we plan to mutate the genomic copies themselves and measure the fitness effect of our mutations when presented in the genome in a single copy or more. Since the deletion of a single copy has little effect over the fitness of the cell (as was demonstrated in the tRNA deletion project that is performed by Zohar Bloom at our lab), it is not certain that these mutation will cause a growth defect that is above the detection level. Ultimately, evolving a strain in which more than half of the 11 tRNA copies are mutated will be the best test to check whether indeed gene conversion maintains the sequence conservation.

7 | Bibliography

1. Innan, H. & Kondrashov, F. The evolution of gene duplications: classifying and distinguishing between models. *Nature reviews. Genetics* **11**, 97–108 (2010).

2. Ohta, T. On the evolution of multigene families. *Theoretical population biology* **23**, 216–40 (1983).

3. Ohta, T. Allelic and nonallelic homology of a supergene family. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 3251–4 (1982).

4. Ohta, T. Genetic variation in small multigene families. *Genetical research* **37**, 133–49 (1981).

5. Sugino, R. P. & Innan, H. Selection for more of the same product as a force to enhance concerted evolution of duplicated genes. *Trends in genetics : TIG* **22**, 642–4 (2006).

6. Innan, H. A two-locus gene conversion model with selection and its application to the human RHCE and RHD genes. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 8793–8 (2003).

7. Mano, S. & Innan, H. The evolutionary rate of duplicated genes under concerted evolution. *Genetics* **180**, 493–505 (2008).

8. Teshima, K. M. & Innan, H. The effect of gene conversion on the divergence between duplicated genes. *Genetics* **166**, 1553–60 (2004).

9. Gangloff, S., Zou, H. & Rothstein, R. Gene conversion plays the major role in controlling the stability of large tandem repeats in yeast. *The EMBO journal* **15**, 1715–1725 (1996).

10. Hillis, D. M., Moritz, C., Porter, C. a & Baker, R. J. Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science (New York, N.Y.)* **251**, 308–10 (1991).

11. San Filippo, J., Sung, P. & Klein, H. Mechanism of eukaryotic homologous recombination. *Annual review of biochemistry* **77**, 229–57 (2008).

12. Krogh, B. O. & Symington, L. S. Recombination proteins in yeast. *Annual review of genetics* **38**, 233–71 (2004).

13. Agmon, N., Pur, S., Liefshitz, B. & Kupiec, M. Analysis of repair mechanism choice during homologous recombination. *Nucleic acids research* **37**, 5081–92 (2009).

14. Pâques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiology and molecular biology reviews : MMBR* **63**, 349–404 (1999).

15. Sugino, R. P. & Innan, H. Estimating the time to the whole-genome duplication and the duration of concerted evolution via gene conversion in yeast. *Genetics* **171**, 63–9 (2005).

16. Osada, N. & Innan, H. Duplication and gene conversion in the Drosophila melanogaster genome. *PLoS genetics* **4**, e1000305 (2008).

17. Ide, S., Miyazaki, T., Maki, H. & Kobayashi, T. Abundance of ribosomal RNA gene copies maintains genome integrity. *Science (New York, N.Y.)* **327**, 693–6 (2010).

18. Eickbush, T. H. & Eickbush, D. G. Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* **175**, 477–85 (2007).

19. Parkin, E. J. & Butlin, R. K. Within- and between-individual sequence variation among ITS1 copies in the meadow grasshopper Chorthippus parallelus indicates frequent intrachromosomal gene conversion. *Molecular biology and evolution* **21**, 1595–601 (2004).

20. Venema, J. & Tollervey, D. Ribosome synthesis in Saccharomyces cerevisiae. *Annual review of genetics* 261–311 (1999).at <http://www.annualreviews.org/doi/pdf/10.1146/annurev.genet.33.1.261>

21. Kobayashi, T. & Ganley, A. R. D. Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats. *Science (New York, N.Y.)* **309**, 1581–4 (2005).

22. Nei, M. & Rooney, A. P. Concerted and birth-and-death evolution of multigene families. *Annual review of genetics* **39**, 121–52 (2005).

23. KOVARIK, A. *et al.* Concerted evolution of 18-5.8-26S rDNA repeats in Nicotiana allotetraploids. *Biological Journal of the Linnean Society* **82**, 615–625 (2004).

24. Schlötterer, C. & Tautz, D. Chromosomal homogeneity of Drosophila ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Current biology : CB* **4**, 777–83 (1994).

25. Petes, T. D. Unequal meiotic recombination within tandem arrays of yeast ribosomal DNA genes. *Cell* **19**, 765–74 (1980).

26. Szostak, J. W. & Wu, R. Unequal crossing over in the ribosomal DNA of Saccharomyces cerevisiae. *Nature* **284**, 426–30 (1980).

27. Ganley, A. R. D. & Kobayashi, T. Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome research* **17**, 184–91 (2007).

28. Santoyo, G. & Romero, D. Gene conversion and concerted evolution in bacterial genomes. *FEMS microbiology reviews* **29**, 169–83 (2005).

29. James, S. a *et al.* Repetitive sequence variation and dynamics in the ribosomal DNA array of Saccharomyces cerevisiae as revealed by whole-genome resequencing. *Genome research* **19**, 626–35 (2009).

30. Stage, D. E. & Eickbush, T. H. Sequence variation within the rRNA gene loci of 12 Drosophila species. *Genome research* **17**, 1888–97 (2007).

31. Hughes, K. & Petersen, R. Apparent recombination or gene conversion in the ribosomal ITS region of a Flammulina (Fungi, Agaricales) hybrid. *Molecular Biology and Evolution* 94–96 (2001).at <http://mbe.oxfordjournals.org/content/18/1/94.short>

32. Wang, S., Zhang, L., Hu, J., Bao, Z. & Liu, Z. Molecular and cellular evidence for biased mitotic gene conversion in hybrid scallop. *BMC evolutionary biology* **10**, 6 (2010).

33. Averbeck, K. T. & Eickbush, T. H. Monitoring the mode and tempo of concerted evolution in the Drosophila melanogaster rDNA locus. *Genetics* **171**, 1837–46 (2005).

34. McTaggart, S. J., Dudycha, J. L., Omilian, A. & Crease, T. J. Rates of recombination in the ribosomal DNA of apomictically propagated Daphnia obtusa lines. *Genetics* **175**, 311–20 (2007).

35. Hopper, A. K., Pai, D. a & Engelke, D. R. Cellular dynamics of tRNAs and their genes. *FEBS letters* **584**, 310–7 (2010).

36. Phizicky, E. M. & Hopper, A. K. tRNA biology charges to the front. *Genes & development* **24**, 1832–60 (2010).

37. Hopper, A. K. & Phizicky, E. M. tRNA transfers to the limelight. *Genes & development* **17**, 162–80 (2003).

38. Barzel, A. & Kupiec, M. Finding a match: how do homologous sequences get together for recombination? *Nature reviews. Genetics* **9**, 27–37 (2008).

39. Duan, Z. *et al.* A three-dimensional model of the yeast genome. *Nature* **465**, 363–367 (2010).

40. Thompson, M., Haeusler, R. a, Good, P. D. & Engelke, D. R. Nucleolar clustering of dispersed tRNA genes. *Science (New York, N.Y.)* **302**, 1399–401 (2003).

41. McFarlane, R. J. & Whitehall, S. K. tRNA genes in eukaryotic genome organization and reorganization. *Cell cycle (Georgetown, Tex.)* **8**, 3102–6 (2009).

42. Pratt-Hyatt, M. J., Kapadia, K. M., Wilson, T. E. & Engelke, D. R. Increased recombination between active tRNA genes. *DNA and cell biology* **25**, 359–64 (2006).

43. De la Loza, M. C. D., Wellinger, R. E. & Aguilera, a Stimulation of direct-repeat recombination by RNA polymerase III transcription. *DNA repair* **8**, 620–6 (2009).

44. Di Rienzi, S. C., Collingwood, D., Raghuraman, M. K. & Brewer, B. J. Fragile genomic sites are associated with origins of replication. *Genome biology and evolution* **1**, 350–63 (2009).

45. Munz, P., Amstutz, H., Kohli, J. & Leupold, U. Recombination between dispersed serine tRNA genes in Schizosaccharomyces pombe. *Nature* **300**, 225–31 (1982).

46. Amstutz, H., Munz, P., Heyer, W. D., Leupoid, U. & Kohli, J. Concerted evolution of tRNA genes: intergenic conversion among three unlinked serine tRNA genes in S. pombe. *Cell* **40**, 879–86 (1985).

47. Kohli, J. *et al.* Interallelic and intergenic conversion in three serine tRNA genes of Schizosaccharomyces pombe. *Cold Spring Harbor symposia on quantitative biology* **49**, 31–40 (1984).

48. Gao, L.-Z. & Innan, H. Very low gene duplication rate in the yeast genome. *Science (New York, N.Y.)* **306**, 1367–70 (2004).

49. Gray, N. K. & Hentze, M. W. Regulation of protein synthesis by mRNA structure. *Molecular biology reports* **19**, 195–200 (1994).

50. Stoletzki, N. & Eyre-Walker, A. Synonymous codon usage in Escherichia coli: selection for translational accuracy. *Molecular biology and evolution* **24**, 374–81 (2007).

51. Varenne, S., Buc, J., Lloubes, R. & Lazdunski, C. Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *Journal of molecular biology* **180**, 549–76 (1984).

52. Tuller, T. *et al.* An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–54 (2010).

53. Dong, H., Nilsson, L. & Kurland, C. G. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *Journal of molecular biology* **260**, 649–63 (1996).

54. Percudani, R., Pavesi, A. & Ottonello, S. Transfer RNA gene redundancy and translational selection in Saccharomyces cerevisiae. *Journal of molecular biology* **268**, 322–30 (1997).

55. Kanaya, S., Yamada, Y., Kudo, Y. & Ikemura, T. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* **238**, 143–55 (1999).

56. Sharp, P. M. & Li, W. H. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic acids research* **15**, 1281–95 (1987).

57. Dos Reis, M., Savva, R. & Wernisch, L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic acids research* **32**, 5036–44 (2004).

58. Novoa, E. M. & Ribas de Pouplana, L. Speeding with control: codon usage, tRNAs, and ribosomes. *Trends in genetics : TIG* **28**, 574–81 (2012).

59. Sharp, P. M. & Li, W. H. Codon usage in regulatory genes in Escherichia coli does not reflect selection for “rare” codons. *Nucleic acids research* **14**, 7737–49 (1986).

60. Duret, L. & Mouchiroud, D. Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 4482–7 (1999).

61. Sharp, P. M. & Li, W. H. An evolutionary perspective on synonymous codon usage in unicellular organisms. *Journal of molecular evolution* **24**, 28–38 (1986).

62. Thanaraj, T. a & Argos, P. Ribosome-mediated translational pause and protein domain organization. *Protein science : a publication of the Protein Society* **5**, 1594–612 (1996).

63. Kramer, G., Boehringer, D., Ban, N. & Bukau, B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nature structural & molecular biology* **16**, 589–97 (2009).

64. Cabrita, L. D., Dobson, C. M. & Christodoulou, J. Protein folding on the ribosome. *Current opinion in structural biology* **20**, 33–45 (2010).

65. Wilke, C. O. & Drummond, D. A. Signatures of protein biophysics in coding sequence evolution. *Current opinion in structural biology* **20**, 385–9 (2010).

66. Withers, M., Wernisch, L. & Dos Reis, M. Archaeology and evolution of transfer RNA genes in the Escherichia coli genome. *RNA (New York, N.Y.)* **12**, 933–42 (2006).

67. Unger, T., Jacobovitch, Y., Dantes, A., Bernheim, R. & Peleg, Y. Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. *Journal of structural biology* **172**, 34–44 (2010).

68. Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H. & Hieter, P. Multifunctional yeast high-copy-number shuttle vectors. *Gene* **110**, 119–22 (1992).

69. Kaganovich, D., Kopito, R. & Frydman, J. Misfolded proteins partition between two distinct quality control compartments. *Nature* **454**, 1088–95 (2008).

70. Weiss, W. A. & Friedberg, E. C. Normal yeast tRNA(CAGGln) can suppress amber codons and is encoded by an essential gene. *Journal of molecular biology* **192**, 725–35 (1986).

71. Chakshusmathi, G., Kim, S. Do, Rubinson, D. A. & Wolin, S. L. A La protein requirement for efficient pre-tRNA folding. *The EMBO journal* **22**, 6562–72 (2003).

72. Johansson, M. J. O., Esberg, A., Huang, B., Björk, G. R. & Byström, A. S. Eukaryotic wobble uridine modifications promote a functionally redundant decoding system. *Molecular and cellular biology* **28**, 3301–12 (2008).

73. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. *The American Naturalist* **138**, 1315 (1991).

74. Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic acids research* **37**, D93–7 (2009).

75. Schmeing, T. M., Voorhees, R. M., Kelley, A. C. & Ramakrishnan, V. How mutations in tRNA distant from the anticodon affect the fidelity of decoding. *Nature structural & molecular biology* **18**, 432–6 (2011).

76. Hillenmeyer, M. E. *et al.* The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science (New York, N.Y.)* **320**, 362–5 (2008).

77. Hillenmeyer, M. E. *et al.* Systematic analysis of genome-wide fitness data in yeast reveals novel gene function and drug action. *Genome biology* **11**, R30 (2010).

78. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337–41 (2009).

79. Schacherer, J., Shapiro, J. a, Ruderfer, D. M. & Kruglyak, L. Comprehensive polymorphism survey elucidates population structure of Saccharomyces cerevisiae. *Nature* **458**, 342–5 (2009).

80. Drummond, D. A. & Wilke, C. O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341–52 (2008).

81. Pechmann, S. & Frydman, J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nature structural & molecular biology* 1–8 (2012).doi:10.1038/nsmb.2466

82. Unternahrer, S., Pridmore, D. & Hinnen, A. A new system for amplifying 2 jam piasmid copy number in Saccharomyces cerevisiae. **5**, 1539–1548 (1991).

83. Ghosh, S. K., Hajra, S., Paek, A. & Jayaram, M. Mechanisms for chromosome and plasmid segregation. *Annual review of biochemistry* **75**, 211–41 (2006).

84. Futcher, A. The 2 Î¼m circle plasmid of Saccharomyces cerevisiae. *Yeast* **4**, (1988).

85. Ling, J., Reynolds, N. & Ibba, M. Aminoacyl-tRNA synthesis and translational quality control. *Annual review of microbiology* **63**, 61–78 (2009).

86. Shimada, a, Nureki, O., Goto, M., Takahashi, S. & Yokoyama, S. Structural and mutational studies of the recognition of the arginine tRNA-specific major identity element, A20, by arginyl-tRNA synthetase. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 13537–42 (2001).

87. Sissler, M., Giegé, R. & Florentz, C. Arginine aminoacylation identity is context-dependent and ensured by alternate recognition sets in the anticodon loop of accepting tRNA transcripts. *The EMBO journal* **15**, 5069–76 (1996).

88. Fender, A. *et al.* A yeast arginine specific tRNA is a remnant aspartate acceptor. *Nucleic acids research* **32**, 5076–86 (2004).

89. Iben, J. R. & Maraia, R. J. tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote. *RNA (New York, N.Y.)* (2012).doi:10.1261/rna.032151.111

90. Symington, L. Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. *Microbiology and molecular biology reviews* **66**, (2002).

91. Breslow, D. K. *et al.* A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nature methods* **5**, 711–8 (2008).

92. Gresham, D. *et al.* The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS genetics* **4**, e1000303 (2008).

93. Widmann, J., Harris, J. & Lozupone, C. Stable tRNA-based phylogenies using only 76 nucleotides. *RNA* 1469–1477 (2010).doi:10.1261/rna.726010.tity

94. Gingold, H. & Pilpel, Y. Determinants of translation efficiency and accuracy. *Molecular systems biology* **7**, 481 (2011).

95. Higgs, P. G. & Ran, W. Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage. *Molecular biology and evolution* **25**, 2279–91 (2008).

96. Gingold, H., Dahan, O. & Pilpel, Y. Dynamic changes in translational efficiency are deduced from codon usage of the transcriptome. *Nucleic acids research* 1–11 (2012).doi:10.1093/nar/gks772

97. Ahn, B. Y., Dornfeld, K. J., Fagrelius, T. J. & Livingston, D. M. Effect of limited homology on gene conversion in a Saccharomyces cerevisiae plasmid recombination system. *Molecular and cellular biology* **8**, 2442–8 (1988).

98. Casola, C., Conant, G. C. & Hahn, M. W. Very low rate of gene conversion in the yeast genome. *Molecular biology and evolution* **29**, 3817–26 (2012).