A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes

Stefanie Mühlhausen[1], Peggy Findeisen[1], Uwe Plessmann[2], Henning Urlaub[2,3] and Martin Kollmar[1§]

[1] Group Systems Biology of Motor Proteins, Department of NMR-based Structural Biology, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

[2] Bioanalytical Mass Spectrometry, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

[3] Bioanalytics Group, Department of Clinical Chemistry, University Medical Center Göttingen, Robert Koch Strasse 40, 37075 Göttingen, Germany

[§]Corresponding author, Tel.: +49-551-2012260, Fax: +49-551-2012202

Email addresses:

    SM: stmu@nmr.mpibpc.mpg.de

    PM: pefi@nmr.mpibpc.mpg.de

    UP: uplessm@gwdg.de

    HU: henning.urlaub@mpibpc.mpg.de

    MK: mako@nmr.mpibpc.mpg.de

Running title

A novel nuclear genetic code alteration in yeasts

Keywords

Genetic code, codon reassignment, tRNA, yeast, proteomics

Abstract

The genetic code is the universal cellular translation table to convert nucleotide into amino acid sequences. Changes to sense codons are expected to be highly detrimental. However, reassignments of single or multiple codons in mitochondria and nuclear genomes demonstrated that the code can evolve. Still, alterations of nuclear genetic codes are extremely rare leaving hypotheses to explain these variations, such as the 'codon capture', the 'genome streamlining' and the 'ambiguous intermediate' theory, in strong debate. Here, we report on a novel sense codon reassignment in *Pachysolen tannophilus*, a yeast related to the Pichiaceae. By generating proteomics data and using tRNA sequence comparisons we show that in *Pachysolen* CUG codons are translated as alanine and not as the universal leucine. The polyphyly of the CUG-decoding tRNAs in yeasts is best explained by a *tRNA loss driven codon reassignment* mechanism. Loss of the CUG-tRNA in the ancient yeast is followed by gradual decrease of respective codons and subsequent codon capture by tRNAs whose anticodon is outside the aminoacyl-tRNA synthetase recognition region. Our hypothesis applies to all nuclear genetic code alterations and provides several testable predictions. We anticipate more codon reassignments to be uncovered in existing and upcoming genome projects.

Introduction

The genetic code determines the translation of nucleotide into amino-acid sequences. Any change in the code altering the meaning of a codon would introduce errors into every translated message and thus be highly detrimental or lethal (Knight et al. 2001a). Therefore, negligible if being optimal or not (Freeland and Hurst 1998; Freeland et al. 2000) the canonical genetic code was long thought to be immutable and termed a "frozen accident" of history (Crick 1968). However, reassignments of single or multiple codons in mitochondria (Knight et al. 2001b) and nuclear genomes (Lozupone et al. 2001; Miranda et al. 2006) demonstrated that the code can evolve (Knight et al. 2001a; Koonin and Novozhilov 2009; Moura et al. 2010). Several codons have been reassigned in independent lineages. Most nuclear code alterations reported so far are stop codon and CUG codon reassignments. Mainly three theories have been proposed to explain reassignments in the genetic code. The *codon capture* hypothesis states that a codon and subsequently its then meaningless cognate tRNA must disappear from the coding genome, before a tRNA with a mutated anticodon appears changing the meaning of the codon (Osawa and Jukes 1989; Osawa et al. 1992). Genome replication bias (GC or AT pressure) is thought to cause codon disappearance. In contrast, the *ambiguous intermediate* hypothesis postulates that either mutant tRNAs, which are charged by more than one aminoacyl-tRNA synthetase, or misreading tRNAs drive genetic code changes (Schultz and Yarus 1994, 1996). The ambiguous codon decoding leads to a gradual codon identity change that is completed upon loss of the wild-type cognate tRNA. The alternative CUG encoding as serine instead of leucine in *Candida* and *Debaryomyces* species (the so-called alternative yeast code; AYCU) has strongly been promoted as example for the *ambiguous intermediate* theory. Stated reasons are that the CUG codon decoding is

- 3 -

ambiguous in many extant *Candida* species (Tuite and Santos 1996; Suzuki et al. 1997), the CUG codon decoding can - at least in part - be converted (Santos et al. 1996; Bezerra et al. 2013), and the origin of the $tRNA^{Ser}_{CAG}$ has been estimated to precede the separation of the *Candida* and *Saccharomyces* genera by about 100 million years (Massey et al. 2003). However, the ambiguous decoding of the CUG triplet in extant "CTG clade" species is caused by slightly inaccurate charging of the $tRNA^{Ser}_{CAG}$ and not by competing tRNAs. Third, the *genome streamlining* hypothesis notes that codon changes are driven by selection to minimize the translation machinery (Andersson and Kurland 1995). This best explains the many codon reassignments and losses in mitochondria. In Saccharomycetaceae mitochondria, for example, ten to twenty-five sense codons are unused and the CUG codons are usually translated as threonine through a $tRNA^{Thr}$, which has evolved from a $tRNA^{His}$ ancestor (Su et al. 2011). In the Eremothecium subbranch, the CUG codons are decoded by alanine, but this modified code did not originate by capture of the CUG codons through an anticodon-mutated $tRNA^{Ala}$ but by switching the acceptor stem identity determinants of the Saccharomycetaceae $tRNA^{Thr}$ from threonine to alanine (Ling et al. 2014).

By analysis of the conservation of amino acid types and CUG codon positions in motor and cytoskeletal proteins we recently showed that these proteins allow to unambiguously assign the standard genetic code or AYCU to yeasts (Mühlhausen and Kollmar 2014). Plotting the assigned code onto the yeast phylogeny demonstrated the AYCU to be polyphyletic with "CTG clade" species and *Pachysolen tannophilus* grouping in different branches. Sequence conservation clearly showed that *Pachysolen* does not encode CUG by leucine. However, *Pachysolen* did not share

any CUG positions with other yeasts. This prompted us to determine the identity of the *Pachysolen* CUG encoding by molecular phylogenetic and proteome analyses.

Results

A new nuclear genetic code in the yeast *Pachysolen tannophilus*

We determined the $tRNA_{CAG}s$ in 60 sequenced yeast species (Table S1) and aligned them against known $tRNA_{CAG}^{Leu}s$ and $tRNA_{CAG}^{Ser}s$. While $tRNA_{CAG}^{Leu}s$ and $tRNA_{CAG}^{Ser}s$ could clearly be classified, the identified *Pachysolen* $tRNA_{CAG}$ sequence was dissimilar to both (Fig. S1). Comparison to all *Candida albicans* cytoplasmic tRNAs suggested close relationship to alanine tRNAs. The alanine identity of the *Pachysolen* $tRNA_{CAG}$ was verified by molecular phylogenetic analyses based on extensive sequence and taxonomic sampling (Fig. 1A, Figs. S1-S5, Table S1). The sequence identity of the *Pachysolen* $tRNA_{CAG}$ to all identified yeast GCN-decoding tRNAs (752 Saccharomycetales clade sequences) is on average 69.3%, ranging from 62.1 to 77.0%. This is slightly below the average 83.6% sequence identity within the GCN-decoding tRNAs (minimum sequence identity is 50.6%) reflecting some sequence divergence beyond the anticodon (Fig. S2). The major general alanine tRNA identity determinants, the discriminator base "A73" and the invariant "G3:U70" wobble base pair as part of the conserved 5' sequence G^1GGC^4 (Musier-Forsyth et al. 1991; Saks et al. 1994; Giegé et al. 1998; Giegé and Eriani 2015), are also present in the *Pachysolen* $tRNA_{CAG}$ (Fig. S1). In contrast, serine tRNAs are characterized by a conserved variable loop sequence, and leucine $tRNA_{NAG}s$ by invariant A35 and G37 nucleotides (Fig. S1) (Saks et al. 1994; Giegé et al. 1998; Giegé and Eriani 2015). However, in *E.coli* not the anticodon sequences

but different tertiary structures seem to be important for discriminating serine and leucine tRNAs (Asahara et al. 1993, 1994). Although identity elements are invariant for the respective codon family box tRNAs, the same elements might be present in other tRNAs where they are located in variable regions. For example, the "G3:U70" wobble base pair, which is usually only found in alanine tRNAs, is also present in Phaffomycetaceae and some Saccharomycetaceae $tRNA_{CAG}^{Leu}s$, although these tRNAs clearly belong to the CTN codon box family tRNAs (Fig. 1A and Figs. S2 to S5) and CUG codons are translated as leucine (Mühlhausen and Kollmar 2014). The presence of G37 in most *Candida* $tRNA_{CAG}^{Ser}s$ also causes partial mischarging by leucine-tRNA synthetases (Suzuki et al. 1997). The anticodon loop of the *Pachysolen* $tRNA_{CAG}$, including the conserved A35 and G37 nucleotides, is identical to leucine tRNAs, while the rest of the sequence is similar to alanine tRNAs (Fig. S1).

Unambiguous translation of the CUG codons by alanine

To verify translation of the CUG codons by alanine we analysed a cytoplasmic extract of laboratory grown Pachysolen by high-resolution mass spectrometry (LC-MS/MS) generating approximately 460,000 high-quality tandem mass spectra. Spectra processing resulted in 27,126 non-redundant peptide matches with a median mass measurement error of about 240 parts per billion (Fig. 1B and Fig. S6). We identified 54% (2,844) of the 5,288 predicted proteins with median protein sequence coverage of approximately 20%, and median numbers of peptides and corresponding PSMs identified per protein of 6 and 9, respectively (Fig. S6). Of the non-redundant peptides, 666 contained sequences with CUG codons fully supported by b- and/or

y-type fragment ions. Almost all of these (99.7%) contain the CUG codons translated as alanine (Figs. 1C and 1D, Fig. S7). None of the identified CUG codons are translated as serine or leucine. The small number of CUG codons translated into other amino acids might be due to differences between our and the sequenced *Pachysolen* strain (Liu et al. 2012), transcription and translation errors, or might represent spurious mischarging of the *Pachysolen* $tRNA_{CAG}^{Ala}$. Substantial mischarging of *Candida* $tRNA_{CAG}^{Ser}$s by leucines has been shown *in vitro* and *in vivo* (Suzuki et al. 1997), but other potential mischargings have never been analysed. We suppose that genetic code changes such as the CUG translation as alanine in *Pachysolen* also remained undetected in the analyses of other sequenced genomes.

History of the CUG-decoding tRNA

To reconstruct the history and origin of all yeast $tRNA_{CAG}$s we performed in-depth phylogenetic analyses of all UCN-decoding tRNAs (serine), GCN-decoding tRNAs (alanine), and CUN-decoding tRNAs (leucine; Figs. S8-S10). These analyses support our previous assumption (Mühlhausen and Kollmar 2014) that all "CTG clade" species' $tRNA_{CAG}$s are serine-tRNAs, and that all Saccharomycetaceae, Phaffomycetaceae, Pichiaceae, and Kuraishia species' $tRNA_{CAG}$s are leucine-tRNAs. Monophyly of the $tRNA_{CAG}^{Ser}$s indicates a common origin in the ancestor of the "CTG clade". The UCN-decoding tRNAs split into two major subbranches, a group of UCG-decoding tRNAs to which the $tRNA_{CAG}^{Ser}$s belong, and a group of UCU-decoding tRNAs. This supports a previous notion (Massey et al. 2003) that the $tRNA_{CAG}^{Ser}$ rather originated from an UCG-tRNA by insertion of an A into the anticodon than from an UCU-tRNA by insertion of a C directly before the

anticodon. Saccharomycetaceae UCA-decoding tRNAs were derived from a UCG-decoding tRNA by duplication and change of anticodon, and Phaffomycetaceae, "CTG clade" species, Pichiaceae, and Kuraishia species' UCA decoding tRNAs have common ancestors with UCU-decoding tRNAs. In contrast to the monophyletic $tRNA^{Ser}_{CAG}$s, the $tRNA^{Leu}_{CAG}$s are polyphyletic and many yeasts contain several $tRNA^{Leu}_{CAG}$s derived from gene duplication of cognate and isoacceptor tRNAs (Fig. 2A). For example, *Yarrowia lipolytica* contains thirteen $tRNA^{Leu}_{CAG}$s, two most probably derived from either an ancestral $tRNA^{Leu}_{CAG}$ or $tRNA^{Leu}_{UAG}$, and eleven derived by gene duplication and mutation from a $tRNA^{Leu}_{AAG}$. The Phaffomycetaceae, Kluyveromyces, *Lachancea kluyveri* and *Eremothecium* $tRNA^{Leu}_{CAG}$s were derived from an ancestral $tRNA^{Leu}_{AAG}$, the *Lachancea thermotolerans* and *Lachancea waltii* $tRNA^{Leu}_{CAG}$s were derived by a recent gene duplication and anticodon mutation from $tRNA^{Leu}_{UAG}$, and the Pichiaceae and Kuraishia species all have derived their $tRNA^{Leu}_{CAG}$s by species-specific events (Fig. 2). The CUN family box tRNAs are split into two major groups, a group of $tRNA^{Leu}_{UAG}$ s and a group of $tRNA^{Leu}_{AAG}$ s. The tRNA$_{GAG}$s are only present in Saccharomycetacea and have been derived from a $tRNA^{Leu}_{AAG}$. The $tRNA^{Leu}_{UAG}$ s form two subgroups, which most probably originated after the split of *Yarrowia lipolytica*. One of the subgroups is restricted to *Yarrowia*, *Pachysolen* and the Pichiacea branches, the other is common to all yeasts and contains the *Saccharomyces* $tRNA^{Leu}_{UAG}$. The latter is unique because it is the only *Saccharomyces cerevisiae* tRNA with an unmodified uridine in the wobble position of the anticodon triplett (Randerath et al. 1979; Johansson and Byström

2005). Unmodified U34s have otherwise only been found in mitochondria, chloroplasts and *Mycoplasma* species. The *Saccharomyces* tRNA$_{UAG}^{Leu}$ is able to translate all six leucine codons (Weissenbach et al. 1977). In contrast, modifications to U34, such as 5'-methoxycarbonylmethyl-2-thiouridine in Gln-, Lys-, and Glu-decoding tRNAs in most if not all prokaryotic and eukaryotic species, often restrict codon recognition to codons ending in A (Johansson et al. 2008; Rezgui et al. 2013). As a consequence, all sequenced yeasts have distinct NNA-decoding and NNG-decoding tRNAs for all respective two-codon families and most four-codon families. However, while all Saccharomycetaceae and Phaffomycetaceae contain this unique tRNA$_{UAG}^{Leu}$, this subbranch tRNA is mutated to tRNA$_{AAG}^{Leu}$ or tRNA$_{CAG}^{Leu}$ in *Pachysolen* and all Pichiaceae (Fig. 2A). The reasons for the unmodified U34 in the *Saccharomyces* tRNA$_{UAG}^{Leu}$ are unknown, but it is tempting to assume that the ancient yeast tRNA$_{UAG}^{Leu}$ contained a modified U34 preventing ambiguous decoding of the CUG codons. If the *Yarrowia-*, *Pachysolen-* and Pichiacea-specific tRNA$_{UAG}^{Leu}$ contained a modified U34, the decoding of the CUG codons in the extant species of these branches would not be ambiguous, which is supported by the unambiguous decoding of the CUG codons by alanine in *Pachysolen*.

Discussion

Assessing models for genetic code evolution

How did such diversity in origin and decoding (leucine versus serine versus alanine) evolve? Already many years ago, several testable predictions of each of the codon reassignment hypotheses have been summarized (Knight et al. 2001a). These,

however, did not include the appearance of a yeast $Ala\text{-}tRNA_{CAG}$. Does the

presence of the $tRNA_{CAG}^{Ala}$ still fit into existing models? According to the *codon*

capture theory, CUG codons need to have disappeared before their reassignment at

the split of the "CTG clade" (Fig. 2). The time frame for codon disappearance is

defined by the splits of Saccharomycetaceae / Phaffomycetaceae and *Pachysolen* /

"CTG-clade" branches (hereafter called *SP* and *PC* branches, respectively). Codon

disappearance must have happened either in the very short time frame between the

split of the *SP* and *PC* branches and the divergence of the "CTG clade", or before

the *SP – PC* split, which would then necessarily include re-appearance of the

$tRNA_{CAG}^{Leu}$ and CUG codons in the *SP* branch. Subsequently, the Ala- and Leu-

$tRNA_{CAG}$s could have captured the still unassigned CUG codon in the *Pachysolen*

and the Pichiacea branches independently from each other. However, disappearance

of an entire codon from a genome by neutral mutations is extremely unlikely to

happen in short time and it is unlikely that genome replication bias caused only

one codon to disappear. The *ambiguous intermediate* theory assumes the presence

of both the cognate $tRNA_{CAG}^{Leu}$ and the new $tRNA_{CAG}^{Ser}$ before the split of the *SP* and

PC branches (Massey et al. 2003). It is highly unlikely that the $tRNA_{CAG}^{Ala}$ had been

present at the same time giving rise to an even more ambiguous decoding by three

tRNAs. More likely seems a scenario including two successive ambiguous

intermediates. In this scenario, a time span with another ambiguous CUG codon

decoding ($tRNA_{CAG}^{Leu}$ competing with $tRNA_{CAG}^{Ala}$) would have followed the split of the

"CTG clade". If the *ambiguous intermediate* theory were true, $tRNA_{CAG}^{Ser}$s and

$tRNA_{CAG}^{Leu}$s in extant species should have been derived from the same ancestral Ser-

and Leu-tRNAs. While this is true for the $tRNA^{Ser}_{CAG}$s of the "CTG clade", the $tRNA^{Leu}_{CAG}$s are polyphyletic and clearly have different origins (Fig. 2, Figs. S8 and S9). In addition, the *ambiguous intermediate* theory would at least require - in case of the more likely separate ambiguous intermediate events - the independent loss of the $tRNA^{Ser}_{CAG}$s in the *SP* and Pichiacea / *Pachysolen* branches (*PP* branch), the loss of the $tRNA^{Ala}_{CAG}$ in the Pichiacea branch as part of the second ambiguous intermediate event, and the independent loss of the $tRNA^{Leu}_{CAG}$s in multiple branches including the branches with altered decoding. The *ambiguous intermediate* theory does also not explain that the same codon became ambiguous in two subbranches of the same taxon in very short time.

The *tRNA loss driven codon reassignment* mechanism

The observed polyphyly of the CAG-tRNAs and the CTN family box reassignments are better described by a *tRNA loss driven codon reassignment* process as follows (Fig. 2): The Saccharomyces type $tRNA^{Leu}_{UAG}$ containing the unmodified U34 appeared by gene duplication of a "normal" $tRNA^{Leu}_{UAG}$ before or after divergence of *Yarrowia* enabling decoding of the CUG codons by wobble base pairing (Crick 1966). As a consequence or as an independent event the ancestor of the *SP* and *PC* clades lost its $tRNA^{Leu}_{CAG}$ by gene loss or mutation. This tRNA loss might not have caused any viability issues because CUG codons could still be decoded as leucine by wobble base pairing (Crick 1966) by the U34-unmodified $tRNA^{Leu}_{UAG}$ .or by, although rather inefficiently, wobble base pairing by the U34-modified $tRNA^{Leu}_{UAG}$.

Subsequently, reduction in translational fidelity (Gromadski et al. 2006) might

have caused the number of CUG codons to gradually decrease. Compared to the most ancient yeast species *Yarrowia*, all analysed yeasts have considerably decreased numbers of CUG codons (Fig. 2) (Mühlhausen and Kollmar 2014). Even the highly GC-rich genomes of *Ogataea parapolymorpha* (Ravin et al. 2013) and *Kuraishia capsulata* (Morales et al. 2013) have less CUG codons suggesting a general strong reduction of CUG codon usage after the divergence of *Yarrowia*. Increased AT pressure has been suggested as reason for reduced GC content at the N3 codon position in "CTG clade" species (Massey et al. 2003), and a similar AT bias is also present in coding genes of *Pachysolen*. However, AT pressure would have applied on all codons and cannot explain the observed reduction in CUG codon usage in all yeasts alone. CUG codon reduction most probably happened by transitions to other leucine codons. Within protein coding regions codon changes within the CTN family box, and also between CTG and TTG, are extremely frequent, demonstrated by the observation that not even closely related *Saccharomyces* species have considerable numbers of conserved leucine codons (Mühlhausen and Kollmar 2014). The unassigned CUG codon could have subsequently been captured by other tRNAs with mutated anticodons. Thus, in the "CTG clade" a mutated Ser-tRNA could capture the CUG codon triggered or supported by the loss of the $tRNA^{Leu}_{UAG}$, which eliminated the possibility of further ambiguous CUG decoding. In the *PP* branch, *Pachysolen* subsequently acquired the $tRNA^{Ala}_{CAG}$ from one of the GCN-decoding tRNAs, and the Pichiaceae and Kuraishia species captured $tRNA^{Leu}_{CAG}$s independently from each other (Figs. 2A and S8). In principle, it is highly unlikely that sense codons can be captured by tRNAs charged with noncognate amino acids because of the high recognition accuracy of the

respective tRNAs by the aminoacyl-tRNA synthetases (Crick 1968; Saks et al. 1994). Aminoacyl-tRNA synthetases (aaRSs) recognize their cognate tRNAs at least at two different regions, most prominently at the discriminator nucleotide "N73" at the acceptor stem, and at the anticodon (Saks et al. 1994; Giegé et al. 1998; Giegé and Eriani 2015). However, the recognition sites of leucine, serine and alanine aaRSs do not include the respective anticodons of leucine, serine, and alanine codons, providing an explanation that only these tRNAs (in addition to any leucine tRNA) could capture the free CUG codon. Changing the anticodons of other tRNAs would disrupts charging by the respective aaRSs and thus prevents capture of noncognate codons. The *codon capture* theory distinguishes itself from the *tRNA loss driven codon reassignment* mainly by the order of events (tRNA loss after or before the reduction of CUG codons, respectively), the degree of CUG codon loss (loss of all codons versus a reduction of CUG codon usage, respectively) and the cause of reduction of CUG codons (AT pressure versus tRNA loss and decreased decoding fidelity, respectively; Fig. 2C). In contrast to the *ambiguous intermediate* hypothesis, the *tRNA loss driven codon reassignment* does not require any ambiguous decoding of the CUG codons. In addition, capture of the CUG codon by different tRNAs is an elemental characteristic of the *tRNA loss driven codon reassignment* and does not need to be split into independent events.

Extending the new hypothesis to the evolution of other codon reassignments

Similar to CUG codon reassignment, our hypothesis can be applied to the reassignment of codons in other nuclear codes such as the stop codons in ciliates and diplomonads (Knight et al. 2001a; Lozupone et al. 2001). Reassignment in these cases might have happened by mutation of the single eukaryotic release factor eRF1

freeing respective stop codons. In contrast to bacteria that often have polycistronic mRNAs, eukaryotic mRNAs are usually monocistronic. Thus, read through of eukaryotic mRNAs does not impose any consequences other than elongation of protein tails by few residues. Examples are the UAA and UAG codons that were captured by single-base mutated Gln- and Glu-tRNAs. Both cognate tRNAs and the tRNAs with stop codons can still be charged since the glutamyl-tRNA synthetase does not discriminate the third position of the anticodon (Nureki et al. 2010). Another example supporting the *tRNA loss driven codon reassignment* hypothesis is *Mycoplasma capricolum*, which lacks a dedicated tRNA for decoding CGG but still contains six CGG codons in its genome (Oba et al. 1991). In contrast to the expectations of the *codon capture* theory, the tRNA is lost but codons are still present, which could, although inefficiently, be translated through wobble base pairing by the CGA-decoding tRNA. For genomes with unused codons such as *Micrococcus luteus*, which lacks AUA and AGA codons (Kano et al. 1993), and also mitrochondria, of which most if not all lack some codons, it is often not obvious whether tRNA loss or codon loss happened first. In genomes, which lack entire codon boxes, codon loss is more likely to have preceded tRNA loss, while tRNA loss most likely happened before codon disappearance of single codons of codon box families.

In mitochondria, codon reassignments have happened rather frequently. However, mitochondrial genomes are extremely small and changes in decoding only affect a few genes. In contrast to most nuclear genomes, where four-codon families require at least two tRNAs, most mitochondria require only one tRNA with a U in the wobble position for all four codons. Therefore, reassignments mainly affect codons in mixed codon boxes such as stop codons,

AUA isoleucine, AGR arginine, and AAA lysine. These reassignments might have happened according to the *tRNA loss driven codon reassignment* hypothesis because all mentioned codons could be translated, although very inefficiently, by another tRNA of the codon family box. Usually, this ambiguous decoding is not competitive and happens only occasionally. However, it preserves the organism's viability after loss of one of the respective tRNAs. For example, the TGA stop codon in mitochondrial genomes can be translated by the tryptophane tRNA$_{CCA}$, which has also been shown to happen at low levels under selective pressure in *Bacillus subtilis* (Lovett et al. 1991). Subsequent to tRNA loss, the inefficient capturing tRNAs are often optimized for decoding both the cognate and the captured codon by mutation (e.g. the tryptophane tRNAs in many mitochondria, where they also decode TGA codons, have a C to U mutation in the wobble position of the anticodon creating a UCU that can pair with both UGA and UGG) and/or by nucleotide modification such as 5'-taurinomethylation of the wobble uridine in mitochondrial methionine tRNA$_{AUR}$ (Suzuki et al. 2011; Watanabe and Yokobori 2011) and lysidine and agmatidine modification of the wobble C in isoleucine tRNA$_{CAU}$ in bacteria and archaea, respectively (Voorhees et al. 2013). Instead of adjusting the tRNA for higher decoding fidelity, the aaRS might become adapted by compensatory mutations to slightly different amino acid identity patterns. This could happen by enabling two different tRNAs be charged by the same aaRS, or by gene duplication and subsequent adaptation of the two aaRSs to the respective tRNAs. An example for the latter case is the decoding of the CUN codons in yeast mitochondria by threonine. The threonine tRNA$_{UAG}$ originated from a histidine tRNA$_{GUG}$ (Su et al. 2011) and requires a dedicated ThrRS for proper charging (Pape et al. 1985). In *Eremothecium*, this tRNA$_{UAG}$ was

mutated to "G3:U70" (Ling et al. 2014) and is accordingly charged by the AlaRS, which is non-discriminative against the anticodon. All these examples support our assumption that only tRNAs charged by aaRS that are non-discriminative against the mutated anticodon bases can subsequently capture free codons. However, there are a few examples for reassignments where a change of a discriminator base should have disfavoured correct charging. For example, UGA is translated as glycine in SR1 bacteria from the human microbiota (Campbell et al. 2013), although both C35 and C36 are usually glycine tRNA identity determinants (Giegé et al. 1998). It is not known, however, whether the SR1 GlyRS is also discriminative against C36. For mitochondrial tRNAs the identity determinants are largely unknown (Salinas-Giegé et al. 2015) and the origins of many of the tRNAs with altered anticodons have never been determined, so that these cases might also fit into the model as soon as more data become available. Nevertheless, the current data suggest a close connection between the tRNAs capturing a free codon and the respective aaRSs being able to correctly charge the cognate and the newly assigned tRNAs.

Predictions based on the tRNA-loss driven codon reassignment theory

Our *tRNA loss driven codon reassignment* hypothesis presents several testable predictions that are mutually exclusive with the *codon capture* and *ambiguous intermediate* theories. We predict identification of (i) further yeast species with $tRNA^{Ser}_{CAG}$ branching before the *SP* and *PC* split or within the group of *Pichia* and *Pachysolen* species, (ii) species with $tRNA^{Ser}_{CAG}$s evolved from serine AGN-decoding tRNAs, and (iii) species with $tRNA^{Leu}_{CAG}$s derived from $tRNA_{NAA}$s. Furthermore, we

anticipate finding further yeasts with $tRNA_{CAG}^{Ala}$ and species without $tRNA_{CAG}^{Leu}$ within the *PP* group. According to the *ambiguous intermediate* theory such findings would need to be treated as further independent ambiguous intermediate events. While this might be theoretically possible, it becomes increasingly unlikely that always the same codon is affected. Based on our assumption that tRNAs can only capture CUG codons if the respective aminoacyl-tRNA synthetases are non-discriminative against the mutated anticodon, we don't expect the CUG codon be captured by other tRNAs than leucine, serine and alanine encoding ones. The predictions of the *tRNA loss driven codon reassignment* model might best be tested by sequencing further yeast species.

Materials and Methods

Growth and lysis of *Pachysolen tannophilus NRRL Y-2460*

Pachysolen tannophilus NRRL Y-2460 was obtained from ATCC (LGC Standards). Cells were grown in YFPD medium at 30°C, harvested by centrifugation (20' at 5,000 x g), and washed and resuspended in lysis buffer (50 mM HEPES pH 6,8, 100 mM KCl). The cells were disrupted by three passages through a French press (20,000 lb/in^2) at 4°C, and intact cells and the cell debris removed by centrifugation (10' at 15,000 x g). The supernatant was subjected to SDS-PAGE gel electrophoresis.

Genome annotation

The *Pachysolen* genome assembly (Liu et al. 2012) has been obtained from NCBI (GenBank accessions CAHV01000001-CAHV01000267). Gene prediction was done

with AUGUSTUS (Stanke and Waack 2003) using the parameter "genemodel=complete", the gene feature set of *Candida albicans*, and the standard codon translation table. The gene prediction resulted in 5,288 predicted proteins, out of which 4,210 contain at least one CUG codon. For mass spectrometry database search the database was multiplied so that each new database contains the CUG codons translated by another amino acid.

Mass spectrometry analysis

SDS-PAGE-separated protein samples were processed as described by Shevchenko *et al.* (Shevchenko et al. 1996) The resuspended peptides in sample loading buffer (2% acetonitrile and 0.1% trifluoroacetic acid) were fractionated and analysed by an online UltiMate 3000 RSLCnano HPLC system (Thermo Fisher Scientific) coupled online to the Q Exactive HF mass spectrometer (Thermo Fisher Scientific). Firstly, the peptides were desalted on a reverse phase C18 precolumn (3 cm long, 100 μm inner diameter, 360μm outer diameter) for 3 minutes. After 3 minutes the precolumn was switched online with the analytical column (30 cm long, 75 μm inner diameter) prepared in-house using ReproSil-Pur C18 AQ 1.9 μm reversed phase resin (Dr. Maisch GmbH). The peptides were separated with a linear gradient of 5–35% buffer (80% acetonitrile and 0.1% formic acid) at a flow rate of 300 nl/min (with back pressure 500 bars) over 90 min gradient time. The precolumn and the column temperature were set to 50°C during the chromatography. The MS data were acquired by scanning the precursors in mass range from 350 to 1600 m/z at a resolution of 70,000 at m/z 200. Top 30 precursor ions were chosen for MS2 by using data-dependent acquisition (DDA) mode at a resolution of 15,000 at m/z 200 with maximum IT of 50 ms. Data analysis and search were performed using MaxQuant v.1.5.2.8 as search engine with 1% FDR

against the *Pachysolen* genome annotation database as annotated above, respectively. Search parameter for searching the precursor and fragment ion masses against the database were as described in Oellerich et al. (Oellerich et al. 2011) except that all peptides shorter than seven amino acids were excluded. To get confidence for the amino acids translated from CUG codons, we determined the observed fragment ions around each CUG-encoded residue. Only amino acids with fragment ions at both sides of the amino acid, which allow the determination of the mass of the respective amino acid, were regarded as supported by the data.

tRNA phylogeny

tRNA genes in 60 sequenced yeast species and four *Schizosaccharomyces* species, which were used as outgroup (Table S1) (Mühlhausen and Kollmar 2014), were identified with tRNAscan (Lowe and Eddy 1997) using standard parameters. All genomes, in which $tRNA_{CAG}s$ were not found by tRNAscan, were searched with BLAST and respective tRNAs reconstructed manually. This especially accounts for the many $tRNA_{CAG}s$ having long introns (up to 287 bp). The intron-free $tRNA_{CAG}s$ were aligned against all *Candida albicans* cytoplasmic tRNAs to identify the closest related tRNA types for in-depth analysis. While the $tRNA_{CAG}^{Leu}s$ and $tRNA_{CAG}^{Ser}s$ were easily identified, the *Pachysolen* $tRNA_{CAG}$ grouped within the *Candida* alanine and valine tRNAs. To finally resolve tRNA codon type relationships and reconstruct $tRNA_{CAG}$ evolution, we increased sequence and taxonomic sampling. Therefore, we randomly selected three to ten homologs from all leucine, serine, and alanine isoacceptor tRNAs from all 60 yeast species, and similar numbers of tRNAs from a selection of valine, phenylalanine, methionine, arginine, isoleucine and threonine

codon types. Identical tRNA$_{CAG}$ sequences from gene duplications were removed resulting in an alignment of 172 tRNA sequences (Fig. S2). To refine the resolution of tRNA relationships within codon family boxes, we manually removed mitochondrial Leu-, Ser-, and Ala-tRNAs from the datasets and performed separate phylogenetic analyses of all NAG-tRNAs (320 leucine isoacceptor tRNAs), NGA-tRNAs (776 serine isoacceptor tRNAs), and NGC-tRNAs (824 alanine isoacceptor tRNAs; Figs. S7-S9). Sequence redundancy was removed using the CD-HIT suite (Li and Godzik 2006) generating reduced alignments of representative sequences of less than 95% identity (80 Leu-tRNAs, 76 Ser-tRNAs, and 70 Ala-tRNAs).

Phylogenetic trees were inferred using Neighbour Joining, Bayesian and Maximum Likelihood based methods as implemented in ClustalW v.2.1 (Chenna et al. 2003), Phase v. 2.0 (Jow et al. 2002; Hudelot et al. 2003) and FastTree v. 2.1.7 (Price et al. 2010), respectively. The most appropriate model of nucleotide substitution was determined with JModelTest v. 2.1.5 (Darriba et al. 2012). Accordingly, FastTree was run with the GTR model for estimating the proportion of invariable sites and the GAMMA model to account for rate heterogeneity. Bootstrapping in ClustalW and FastTree was performed with 1,000 replicates. Phase was used with a mixed model, the REV-Γ model for the loops and the RNA7D-Γ model for the stem regions, which were given by a manually generated consensus tRNA secondary structure. Phase was run with 750,000 burnin and 3,000,000 sampling iterations, and a sampling period of 150 cycles. The phylogenetic network was generated with SplitsTree v.4.1.3.1 (Huson and Bryant 2006) using the Neighbor-Net method to identify alternative splits.

Acknowledgments

Author contributions

MK initiated the study. SM performed MS data and phylogenetic analyses. PF prepared experimental samples. UP performed MS experiments. HU was involved in MS data interpretation. MK assembled, aligned and analysed tRNA sequences. SM and MK wrote the manuscript.

Disclosure declaration

None declared.

Figure legends

Figure 1

***Pachysolen* translates CUG codons with alanine instead of the universal leucine.**

a) Unrooted phylogenetic network of 172 tRNA sequences as generated using the Neighbour-Net method as implemented in SplitsTree v4.1.3.1. Methionine, isoleucine, arginine, and threonine tRNAs were included as outgroup. Leu-, Ser-, and Ala- $tRNA_{CAG}$s are highlighted in dark green, blue, and red, respectively. b) Distribution of the mass errors of randomly selected 10% of all peptides (grey dots),

peptides with CUG translated as alanine (red), and peptides with CUG translated as other amino acids (blue). c) Number and distribution of observed CUG translations in all peptide spectrum matches from MS/MS analysis. d) Representative MS/MS spectrum of a peptide containing CUG translated as alanine. The peptide sequence is shown below the spectrum, with the annotation of the identified matched N-terminal fragment ions (b-type ions) in blue and the C-terminal fragment ions (y-type ions) in red. Only major identified peaks are labeled for clarity (full annotation is shown in Fig. S7).

Figure 2

The mechanism of CUG codon reassignment.

a) Decoding and origin of the $tRNA_{CAG}s$ were plotted onto a time-calibrated yeast phylogeny adapted from (Mühlhausen and Kollmar 2014) using the same colour scheme as in Fig. 1A. Coloured lines denote the presence of a $tRNA_{CAG}$ with the colour representing the origin of the respective $tRNA_{CAG}$. For example, *Yarrowia* contains thirteen $tRNA_{CAG}^{Leu}s$, of which two most probably derived from either an ancestral $tRNA_{CAG}^{Leu}$ or $tRNA_{UAG}^{Leu}$ (mantis green line) and eleven most probably derived from an ancestral $tRNA_{AAG}^{Leu}$ (yellow green line), and *Pachysolen* contains a $tRNA_{CAG}^{Ala}$ (red line). Branches with similar decoding schemes have been collapsed. Filled and empty circles at branches denote the presence and absence, respectively, of the CUN family box tRNAs, and arrows indicate tRNA gene duplications followed by anticodon mutations. The percentages of CUG codons per species and taxon were derived from (Mühlhausen and Kollmar 2014). b) The four CUN codons can be decoded by several combinations of tRNAs using standard and wobble base

pairing. c) The scheme contrasts the presence of $tRNA_{CAG}s$ and number of CUG

codons according to the *tRNA loss driven codon reassignment* hypothesis with

assumptions based on the *codon capture* and *ambiguous intermediate* theories.

References

Andersson SG, Kurland CG. 1995. Genomic evolution drives the evolution of the translation system. *Biochem Cell Biol Biochim Biol Cell* **73**: 775–787.

Asahara H, Himeno H, Tamura K, Nameki N, Hasegawa T, Shimizu M. 1993. Discrimination among E. coli tRNAs with a long variable arm. *Nucleic Acids Symp Ser* 207–208.

Asahara H, Himeno H, Tamura K, Nameki N, Hasegawa T, Shimizu M. 1994. Escherichia coli seryl-tRNA synthetase recognizes tRNA(Ser) by its characteristic tertiary structure. *J Mol Biol* **236**: 738–748.

Bezerra AR, Simões J, Lee W, Rung J, Weil T, Gut IG, Gut M, Bayés M, Rizzetto L, Cavalieri D, et al. 2013. Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification. *Proc Natl Acad Sci U S A* **110**: 11079–11084.

Campbell JH, O'Donoghue P, Campbell AG, Schwientek P, Sczyrba A, Woyke T, Söll D, Podar M. 2013. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Proc Natl Acad Sci U S A* **110**: 5540–5545.

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**: 3497–500.

Crick FH. 1966. Codon--anticodon pairing: the wobble hypothesis. *J Mol Biol* **19**: 548–555.

Crick FH. 1968. The origin of the genetic code. *J Mol Biol* **38**: 367–379.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**: 772.

Freeland SJ, Hurst LD. 1998. The genetic code is one in a million. *J Mol Evol* **47**: 238–248.

Freeland SJ, Knight RD, Landweber LF, Hurst LD. 2000. Early fixation of an optimal genetic code. *Mol Biol Evol* **17**: 511–518.

Giegé R, Eriani G. 2015. Transfer RNA Recognition and Aminoacylation by Synthetases. In *eLS*, John Wiley & Sons, Ltd

http://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0000531.pub3/abstract (Accessed July 21, 2015).

Giegé R, Sissler M, Florentz C. 1998. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res* **26**: 5017–5035.

Gromadski KB, Daviter T, Rodnina MV. 2006. A uniform response to mismatches in codon-anticodon complexes ensures ribosomal fidelity. *Mol Cell* **21**: 369–377.

Hudelot C, Gowri-Shankar V, Jow H, Rattray M, Higgs PG. 2003. RNA-based phylogenetic methods: application to mammalian mitochondrial RNA sequences. *Mol Phylogenet Evol* **28**: 241–252.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254–267.

Johansson MJO, Byström AS. 2005. Transfer RNA modifications and modifying enzymes in Saccharomyces cerevisiae. In *Fine-Tuning of RNA Functions by Modification and Editing* (ed. H. Grosjean), *Topics in Current Genetics*, pp. 87–120, Springer Berlin Heidelberg http://link.springer.com/chapter/10.1007/b105814 (Accessed July 20, 2015).

Johansson MJO, Esberg A, Huang B, Björk GR, Byström AS. 2008. Eukaryotic wobble uridine modifications promote a functionally redundant decoding system. *Mol Cell Biol* **28**: 3301–3312.

Jow H, Hudelot C, Rattray M, Higgs PG. 2002. Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. *Mol Biol Evol* **19**: 1591–1601.

Kano A, Ohama T, Abe R, Osawa S. 1993. Unassigned or nonsense codons in Micrococcus luteus. *J Mol Biol* **230**: 51–56.

Knight RD, Freeland SJ, Landweber LF. 2001a. Rewiring the keyboard: evolvability of the genetic code. *Nat Rev Genet* **2**: 49–58.

Knight RD, Landweber LF, Yarus M. 2001b. How mitochondria redefine the code. *J Mol Evol* **53**: 299–313.

Koonin EV, Novozhilov AS. 2009. Origin and evolution of the genetic code: the universal enigma. *IUBMB Life* **61**: 99–111.

Ling J, Daoud R, Lajoie MJ, Church GM, Söll D, Lang BF. 2014. Natural reassignment of CUU and CUA sense codons to alanine in Ashbya mitochondria. *Nucleic Acids Res* **42**: 499–508.

Liu X, Kaas RS, Jensen PR, Workman M. 2012. Draft genome sequence of the yeast Pachysolen tannophilus CBS 4044/NRRL Y-2460. *Eukaryot Cell* **11**: 827.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.

Lovett PS, Ambulos NP, Mulbry W, Noguchi N, Rogers EJ. 1991. UGA can be decoded as tryptophan at low efficiency in Bacillus subtilis. *J Bacteriol* **173**: 1810–1812.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Res* **25**: 0955–964.

Lozupone CA, Knight RD, Landweber LF. 2001. The molecular basis of nuclear genetic code change in ciliates. *Curr Biol CB* **11**: 65–74.

Massey SE, Moura G, Beltrão P, Almeida R, Garey JR, Tuite MF, Santos MAS. 2003. Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in Candida spp. *Genome Res* **13**: 544–557.

Miranda I, Silva R, Santos MAS. 2006. Evolution of the genetic code in yeasts. *Yeast* **23**: 203–213.

Morales L, Noel B, Porcel B, Marcet-Houben M, Hullo M-F, Sacerdot C, Tekaia F, Leh-Louis V, Despons L, Khanna V, et al. 2013. Complete DNA sequence of Kuraishia capsulata illustrates novel genomic features among budding yeasts (Saccharomycotina). *Genome Biol Evol* **5**: 2524–2539.

Moura GR, Paredes JA, Santos MAS. 2010. Development of the genetic code: Insights from a fungal codon reassignment. *FEBS Lett* **584**: 334–341.

Mühlhausen S, Kollmar M. 2014. Molecular Phylogeny of Sequenced Saccharomycetes Reveals Polyphyly of the Alternative Yeast Codon Usage. *Genome Biol Evol* evu152.

Musier-Forsyth K, Usman N, Scaringe S, Doudna J, Green R, Schimmel P. 1991. Specificity for aminoacylation of an RNA helix: an unpaired, exocyclic amino group in the minor groove. *Science* **253**: 784–786.

Nureki O, O'Donoghue P, Watanabe N, Ohmori A, Oshikane H, Araiso Y, Sheppard K, Söll D, Ishitani R. 2010. Structure of an archaeal non-discriminating glutamyl-tRNA synthetase: a missing link in the evolution of Gln-tRNAGln formation. *Nucleic Acids Res* **38**: 7286–7297.

Oba T, Andachi Y, Muto A, Osawa S. 1991. CGG: an unassigned or nonsense codon in Mycoplasma capricolum. *Proc Natl Acad Sci U S A* **88**: 921–925.

Oellerich T, Bremes V, Neumann K, Bohnenberger H, Dittmann K, Hsiao H-H, Engelke M, Schnyder T, Batista FD, Urlaub H, et al. 2011. The B-cell antigen receptor signals through a preformed transducer module of SLP65 and CIN85. *EMBO J* **30**: 3620–3634.

Osawa S, Jukes TH. 1989. Codon reassignment (codon capture) in evolution. *J Mol Evol* **28**: 271–278.

Osawa S, Jukes TH, Watanabe K, Muto A. 1992. Recent evidence for evolution of the genetic code. *Microbiol Rev* **56**: 229–264.

Pape LK, Koerner TJ, Tzagoloff A. 1985. Characterization of a yeast nuclear gene (MST1) coding for the mitochondrial threonyl-tRNA1 synthetase. *J Biol Chem* **260**: 15362–15370.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. *PloS One* **5**: e9490.

Randerath E, Gupta RC, Chia LL, Chang SH, Randerath K. 1979. Yeast tRNA Leu UAG. Purification, properties and determination of the nucleotide sequence by radioactive derivative methods. *Eur J Biochem FEBS* **93**: 79–94.

Ravin NV, Eldarov MA, Kadnikov VV, Beletsky AV, Schneider J, Mardanova ES, Smekalova EM, Zvereva MI, Dontsova OA, Mardanov AV, et al. 2013. Genome sequence and analysis of methylotrophic yeast Hansenula polymorpha DL1. *BMC Genomics* **14**: 837.

Rezgui VAN, Tyagi K, Ranjan N, Konevega AL, Mittelstaet J, Rodnina MV, Peter M, Pedrioli PGA. 2013. tRNA tKUUU, tQUUG, and tEUUC wobble position modifications fine-tune protein translation by promoting ribosome A-site binding. *Proc Natl Acad Sci U S A* **110**: 12289–12294.

Saks ME, Sampson JR, Abelson JN. 1994. The transfer RNA identity problem: a search for rules. *Science* **263**: 191–197.

Salinas-Giegé T, Giegé R, Giegé P. 2015. tRNA biology in mitochondria. *Int J Mol Sci* **16**: 4518–4559.

Santos MA, Perreau VM, Tuite MF. 1996. Transfer RNA structural change is a key element in the reassignment of the CUG codon in Candida albicans. *EMBO J* **15**: 5060–5068.

Schultz DW, Yarus M. 1996. On malleability in the genetic code. *J Mol Evol* **42**: 597–601.

Schultz DW, Yarus M. 1994. Transfer RNA mutation and the malleability of the genetic code. *J Mol Biol* **235**: 1377–1380.

Shevchenko A, Wilm M, Vorm O, Jensen ON, Podtelejnikov AV, Neubauer G, Shevchenko A, Mortensen P, Mann M. 1996. A strategy for identifying gel-separated proteins in sequence databases by MS alone. *Biochem Soc Trans* **24**: 893–896.

Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**: ii215 – 225.

Su D, Lieberman A, Lang BF, Simonovic M, Söll D, Ling J. 2011. An unusual tRNAThr derived from tRNAHis reassigns in yeast mitochondria the CUN codons to threonine. *Nucleic Acids Res* **39**: 4866–4874.

Suzuki T, Miyauchi K, Suzuki T, Yokobori S, Shigi N, Kondow A, Takeuchi N, Yamagishi A, Watanabe K. 2011. Taurine-containing uridine modifications in

tRNA anticodons are required to decipher non-universal genetic codes in ascidian mitochondria. *J Biol Chem* **286**: 35494–35498.

Suzuki T, Ueda T, Watanabe K. 1997. The "polysemous" codon--a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. *EMBO J* **16**: 1122–1134.

Tuite MF, Santos MA. 1996. Codon reassignment in Candida species: an evolutionary conundrum. *Biochimie* **78**: 993–999.

Voorhees RM, Mandal D, Neubauer C, Köhrer C, RajBhandary UL, Ramakrishnan V. 2013. The structural basis for specific decoding of AUA by isoleucine tRNA on the ribosome. *Nat Struct Mol Biol* **20**: 641–643.

Watanabe K, Yokobori S-I. 2011. tRNA Modification and Genetic Code Variations in Animal Mitochondria. *J Nucleic Acids* **2011**: 623095.

Weissenbach J, Dirheimer G, Falcoff R, Sanceau J, Falcoff E. 1977. Yeast tRNALeu (anticodon U--A--G) translates all six leucine codons in extracts from interferon treated cells. *FEBS Lett* **82**: 71–76.