Yeast as a model organism for studying the evolution of non-standard genetic codes

Raquel M. Silva, Isabel Miranda, Gabriela Moura and Manuel A. S. Santos

Date received: 26th December 2003

Raquel M. Silva
is a researcher in the Functional Evolutionary Genomics Laboratory of the Centre for Cell Biology at the University of Aveiro, Portugal.

Isabel Miranda
is a researcher in the Functional Evolutionary Genomics Laboratory of the Centre for Cell Biology at the University of Aveiro, Portugal.

Gabriela Moura
is a senior researcher in the Functional Evolutionary Genomics Laboratory of the Centre for Cell Biology at the University of Aveiro, Portugal.

Manuel A. S. Santos
is an auxiliary professor in Molecular Biology at the Centre for Cell Biology of the University of Aveiro, Portugal.

Keywords: *genetic code, molecular evolution, comparative genomics, proteomics, ribosome, mRNA decoding*

Abstract

During the last 30 years, a number of alterations to the standard genetic code have been uncovered both in prokaryotes and eukaryotic nuclear and mitochondrial genomes. But, the study of the evolutionary pathways and molecular mechanisms of codon identity redefinition has been largely ignored due to the assumption that non-standard genetic codes can only evolve through neutral evolutionary mechanisms and that they have no functional significance. The recent discovery of a genetic code change in the genus *Candida* that evolved through an ambiguous messenger RNA decoding mechanism is bringing that naive assumption to an abrupt end by showing, in a rather dramatic way, that genetic code changes have profound physiological and evolutionary consequences for the species that redefine codon identity. In this paper, the recent data on the evolution of the *Candida* genetic code are reviewed and an experimental framework based on forced evolution, molecular genetics and comparative and functional genomics methodologies is put forward for the study of non-standard genetic codes and genetic code ambiguity in general. Additionally, the importance of using *Saccharomyces cerevisiae* as a model organism for elucidating the evolutionary pathway of the *Candida* and other genetic code changes is emphasised.

INTRODUCTION

The discovery of several genetic code changes in bacteria and eukaryotic nuclear and mitochondrial systems (Figure 1)[1–5] prompted the development of two theories to explain their evolution — namely, the 'Codon Capture' and the 'Ambiguous Intermediate' theories.[6–11]

The 'Codon Capture' theory proposes a neutral mechanism for the evolution of genetic code changes and was developed from the observation that biased GC content has a strong impact on codon usage and, in extreme cases, can drive codons to extinction (Figure 2A,B).[6–8] This theory is supported by the disappearance of the CGG codon in *Mycoplasma capricolum* (25 per cent of genome is GC) and the AGA and AUA codons in *Micrococcus luteus* (75 per cent of genome is GC) (Figure 2B).[7,12] The theory postulates that, once a codon disappears from the genome, it can be reintroduced by genetic drift. In this case, a misreading transfer RNA (tRNA) from a non-cognate codon family captures the reintroduced codon, thus changing its identity. At this stage, the existence of a misreading tRNA is critical for identity redefinition, otherwise the codon remains unassigned (ie does not code for any amino acid). The usage of the reintroduced codon having a new identity — due to its capture by a non-cognate misreading tRNA — can then increase over time.[6–8] This neutral theory excludes any function for genetic code changes and postulates, in a rather elegant and robust manner, that genetic code changes result from biased genome GC content arising from mutations in DNA polymerases or DNA repair systems (Figure 2A).[7]

The discovery of additional genetic code changes involving both sense and non-sense codons has questioned the 'Codon Capture' theory because it became apparent that certain codons change identity without vanishing from

Manuel A. S. Santos,
Centre for Cell Biology,
Department of Biology,
University of Aveiro,
3810-193 Aveiro, Portugal

Tel: +351 234 370771
Fax: +351 234 426408
E-mail: msantos@bio.ua.pt

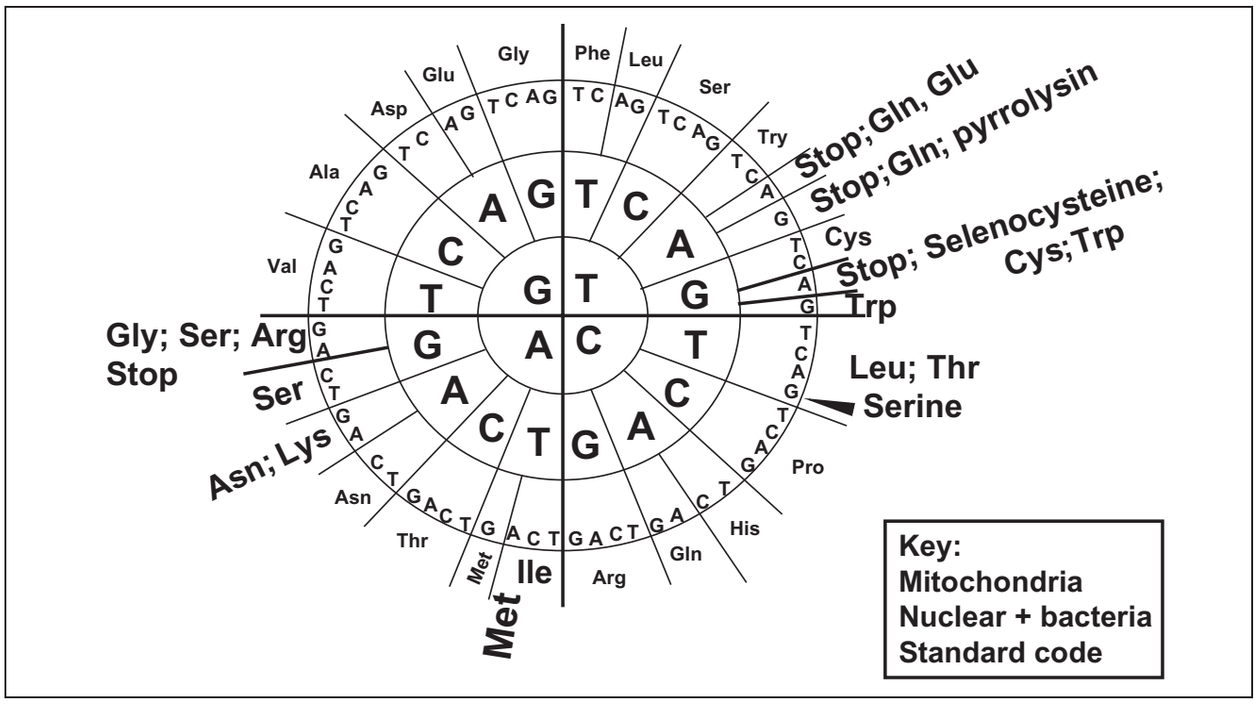


Figure 1: Diversity of the genetic code. To date, a number of genetic code changes have been uncovered in eubacteria, archaea and in eukaryotic nuclear and mitochondrial genomes. In mitochondrial systems, both sense and non-sense codons can change identity; conversely, in prokaryotic and eukaryotic nuclear systems, only stop codons change identity, the exception being the decoding of the leucine–CUG codon as serine in *Candida* species. The prokaryotic and eukaryotic nuclear genetic code changes are a subset of the mitochondrial ones, indicating that there are some codon sets that are more prone to changing their identity than others. In particular, codons starting with T or A change their identity rather often, while codons starting with C only change their identity in yeast mitochondria and in *Candida*, where the leucine CUN codon family is decoded as threonine and the CUG codon is decoded as serine, respectively. Codons starting with G are apparently resistant to identity changes, suggesting that the strength of codon–anticodon interaction at the first codon position is an important determinant of the evolution of non-standard genetic codes

Genetic code changes can be mediated via biased genome GC pressure and ambiguous codon decoding

the genome.[5] This is supported by the observation that GC content along the genome is not evenly distributed[13] and, consequently, GC pressure at the third codon position — the one that would allow codon disappearance without negative impact on protein sequence — is not the same in all parts of the genome.[13] Additionally, some codons that change identity do so in a manner that is unrelated to genome GC content.[5,14] So, how do codons change their identity without disappearing from the genome? The most probable hypothesis is that mutant tRNAs with double identity (recognised by more than one aminoacyl–tRNA synthetase) or with expanded decoding properties (altered anticodons or unusual tertiary structures) provide a route for driving genetic code changes through an ambiguous codon decoding mechanism.[14,15] The existence of natural non-sense suppressor tRNAs — natural or mutant tRNAs that decode one of the three termination codons UAA, UAG or UGA — combined with the observation that stop codons are often reassigned both in eukaryotic cytoplasmic and mitochondrial translation systems and also in several bacteria (Figure 1) provide strong support for the hypothesis that codon identity redefinition can be mediated through mechanisms involving ambiguous codon decoding (Figure 3A,B).[16,17]

The above observations prompted Yarus and Schultz to propose the 'Ambiguous Intermediate' theory for

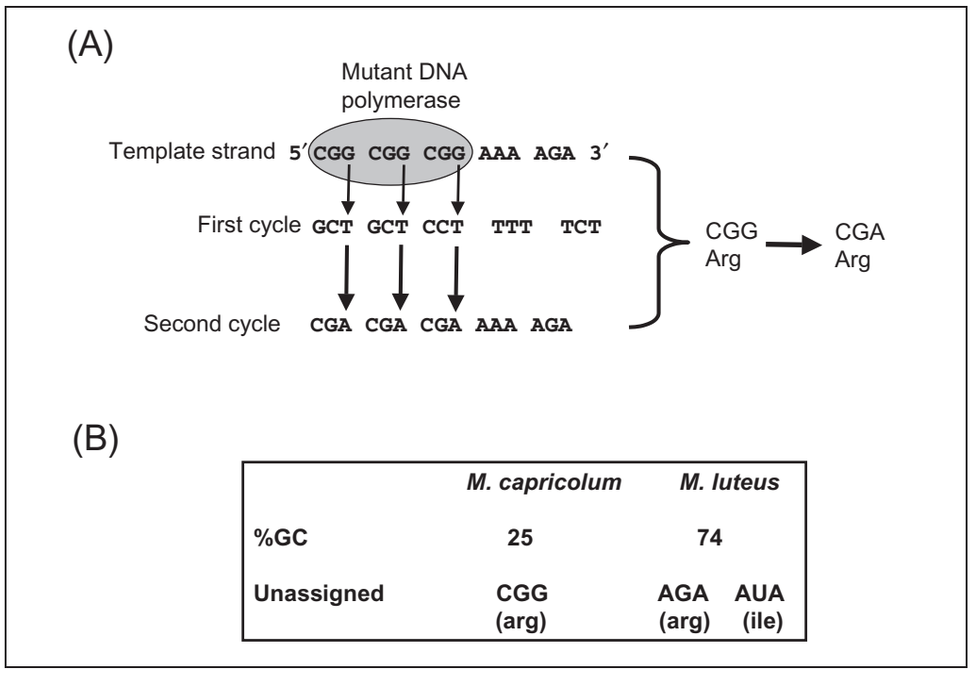


Figure 2: Codon unassignment and capture through biased genome GC pressure. (A) Strong GC pressure arising from mutations in DNA polymerases and/or DNA repair systems alters codon usage significantly. In extreme cases, biased GC pressure may lead to codon disappearance from the genome. (B) The disappearance of the arginine AGA and the isoleucine AUA codons from the genome of *Micrococcus luteus* (74 per cent GC) and the arginine CGG codon from the genome of *Mycoplasma capricolum* (25 per cent GC) provides strong support for this mechanism.[7] A similar mechanism is apparently in action in the AT-rich mitochondrial genomes which display a rather high genetic code diversity. In these cases, degeneracy of the genetic code works as a buffer permitting codon disappearance without changing protein sequence. Once a codon has disappeared from the genome, it can be reintroduced by genetic drift, allowing for redefinition of the identity of the newly introduced codon if a tRNA from a non-cognate codon family misreads it. The existence of a misreading tRNA (wild-type or mutant) is critical to ensure decoding of the reintroduced codon — if the codon is not decoded, it will stall the ribosome during mRNA translation, blocking protein synthesis. The misreading tRNA can then acquire a cognate anticodon for the 'new' codon which will allow for its efficient decoding during mRNA translation and full capture of the reintroduced codon[6–8]

Biased genome replication due to mutations in DNA polymerases alter codon usage and in extreme cases may force codon disappearance

the evolution of non-standard genetic codes.[9–11] This is a non-neutral theory that postulates that ambiguous codon decoding provides a critical initial step for gradual codon identity change (Figure 3A,B). In this scenario, wild-type or mutant misreading tRNAs have the potential to decode non-cognate codons and, in extreme cases, take over their decoding from their cognate tRNAs; however, complete codon identity redefinition can only be accomplished if the wild-type cognate tRNA decoders disappear from the genome by negative selection.[9,10] Intriguingly, this rather puzzling mechanism that destabilises the proteome and reduces fitness has been given strong support by the discovery that certain extant *Candida* species redefined the identity of the CUG codon from leucine to serine through an ambiguous decoding mechanism.[18–22] Remarkably, in *C. zeylanoides*, the novel Ser–tRNA (Ser–$tRNA_{CAG}$) that mediates this genetic code change is still ambiguous because it is charged with both leucine (3–5 per cent) and serine

The identity of the leucine CUG codon has been redefined to serine in *Candida* spp. but remains ambiguous in some species

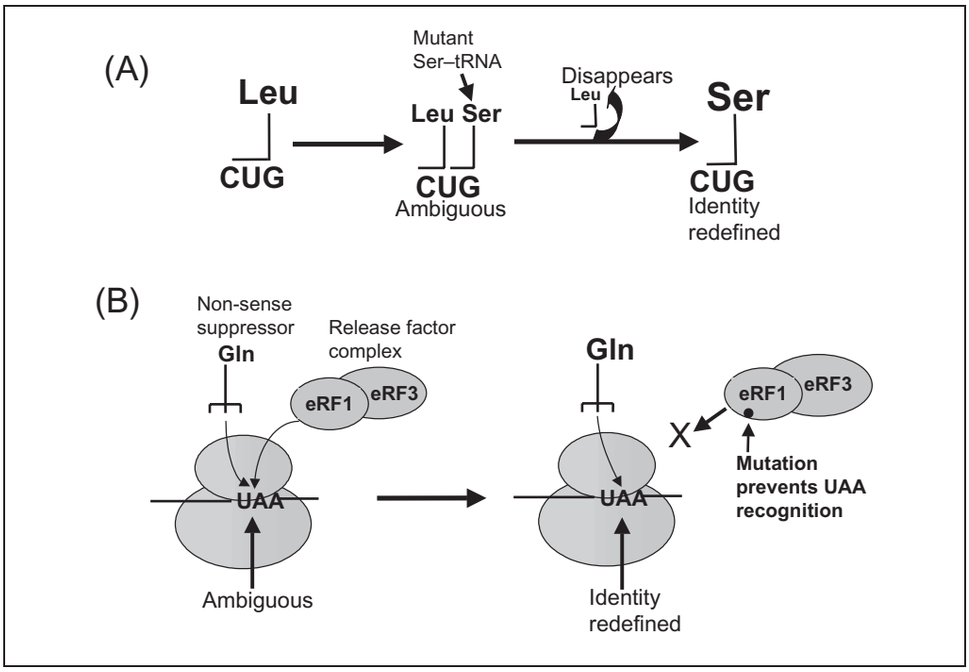


Figure 3: Mechanisms of codon identity redefinition through genetic code ambiguity. (A) Sense codon identity redefinition. The identity of sense codons can be redefined gradually through a molecular mechanism that involves ambiguous decoding during an intermediary stage of codon identity redefinition.[9–11] This ambiguity can arise through competition between a cognate tRNA and a non-cognate misreading tRNA or through mischarging of a cognate tRNA by an aminoacyl–tRNA synthetase belonging to a different tRNA isoacceptor family. The latter is particularly relevant because tRNA charging *in vivo* depends on the correct balance between aminoacyl–tRNA synthetases and tRNAs.[18] The identity of the codon is completely redefined on disappearance of the tRNA that decoded the codon according to the standard genetic code rules. The diagram shows the redefinition of the identity of the leucine CUG codon to serine in the genus *Candida*. (B) Non-sense codon identity redefinition. The identity of non-sense codons can also be redefined gradually through ambiguous decoding involving an intermediary stage characterised by competition between non-sense suppressor tRNAs and the release factor complex, which recognises the stop codons UAA, UAG or UGA and terminates polypeptide chain elongation during mRNA translation. The diagram shows the redefinition of the identity of the UAA stop codon to glutamine that occurs in several species of ciliates, namely *Tetrahymena thermophila*. For complete identity change of the UAA stop codon, the anticodon (5′-UUG-3′) of the Gln–tRNA mutates creating an anticodon (5′-UUA-3′) that is cognate for the UAA codon. Additionally, eRF1 loses its ability to recognise the UAA stop codon (its ability to recognise the other stop codons is not altered), thus ensuring efficient decoding of the UAA codon as glutamine. The same sequence of events is valid for altering the identity of the other two non-sense codons, which occurs in mitochondria, bacteria and ciliates[5,16]

Codon decoding ambiguity is an important mechanism of codon sense and non-sense codon identity redefinition

(95–97 per cent),[23] thus providing direct evidence for the hypothesis that CUG identity redefinition evolved through an ambiguous decoding mechanism.[23] The functional role of this codon decoding ambiguity still present in some *Candida* species is not yet understood.[15,24]

COMPARATIVE AND EVOLUTIONARY GENOMICS

An important characteristic of genetic code changes is that they evolve from the standard genetic code and do not represent ancestral lineages of alternative genetic codes that existed in the RNA

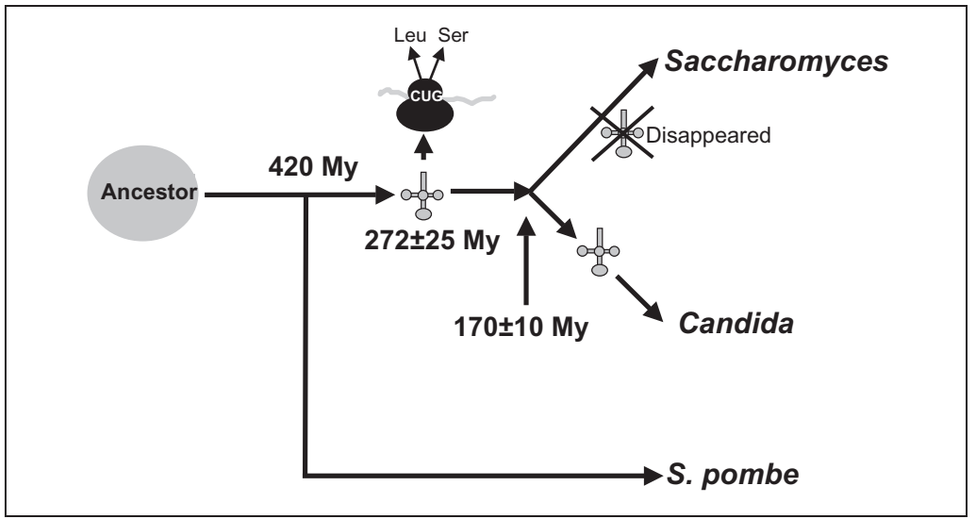
Non-standard genetic codes evolved recently from the standard genetic code

world.[5] Therefore, timing the evolutionary pathway of genetic code alterations is of fundamental importance in better understanding the mechanisms and forces driving change. This assumes particular relevance in the context of genetic code changes that evolve through ambiguous codon decoding as these evolve gradually and introduce significant proteome destabilisation whose relevance can only be appreciated on an evolutionary time-scale.

By using ribosomal RNA (rRNA) and tRNA molecular phylogeny approaches, Massey and colleagues[25] have shed the very first light onto the time-scale of the *Candida* CUG identity redefinition from leucine to serine. These authors have shown that high-level serine–CUG decoding is approximately 171±27 million years (My) old; however, the ser–$tRNA_{CAG}$ that decodes it as serine originated at least 272±25 My ago.

These molecular phylogeny studies showed that the probability of the change in CUG identity and the ser–$tRNA_{CAG}$ appearing at the same time — within 10 My of each other — is extremely low ($p < 0.0006$). Interestingly, the same molecular phylogeny methodology indicated that the geni *Candida* and *Saccharomyces* separated from each other 178±19 My ago, implying that the CUG codon was highly ambiguous in the ancestor of these yeasts for approximately 100 My (Figure 4).[25,26]

During the early stages of CUG identity redefinition, decoding ambiguity arose via competition between the leu–$tRNA_{CAG}$ and the new ser–$tRNA_{CAG}$ for the decoding of the CUG codon at the ribosome A site during mRNA translation. The consequences of this codon decoding ambiguity on an evolutionary time-scale raise two fundamental biological questions:



Redefinition of the identity of the CUG codon from leucine to serine in *Candida* spp. is approximately 272 My old

Figure 4: The evolutionary pathway of CUG reassignment in *Candida*. In *Candida*, the identity of the leucine–CUG codon has been redefined to serine by a novel serine–tRNA (ser–$tRNA_{CAG}$) that acquired a 5′-CAG-3′ anticodon allowing it to decode the CUG codon as serine. Molecular phylogeny studies using both rRNA and tRNA sequences showed that the ser–$tRNA_{CAG}$ appeared approximately 272 million years (My) ago, while the genera *Candida* and *Saccharomyces* separated from each other 170 My.[25] This reassignment pathway suggests that the *Saccharomyces* ancestor was ambiguous for at least 100 My and that the genus *Schizosaccharomyces* was not affected by this genetic code change.[25] The ser–$tRNA_{CAG}$ was maintained in the lineage that originated the genus *Candida* but was lost in the lineage leading to the genus *Saccharomyces*. Complete CUG reassignment required the disappearance of the cognate leu–tRNA[25]

- Does decoding ambiguity affect the usage of the ambiguous codons?
- What are the physiological and evolutionary consequences of 'living' with an unstable proteome on an evolutionary time-scale?

Considering that codon identity redefinition has the potential to affect the entire set of genes of a genome and destabilise all, or almost all, proteins encoded by that genome; these questions can only be tackled using comparative (Figure 5) and functional genomics approaches (Figures 6–8). For this, the powerful genomics and proteomics tools available to yeast researchers and the availability of the complete genome sequences of several yeast species — namely, *S. cerevisiae*, *C. albicans* and *Schizosaccharomyces pombe* — are of paramount importance.

Comparative genomics allows for tracing the fate of ambiguous CUG codons in *Candida* spp.

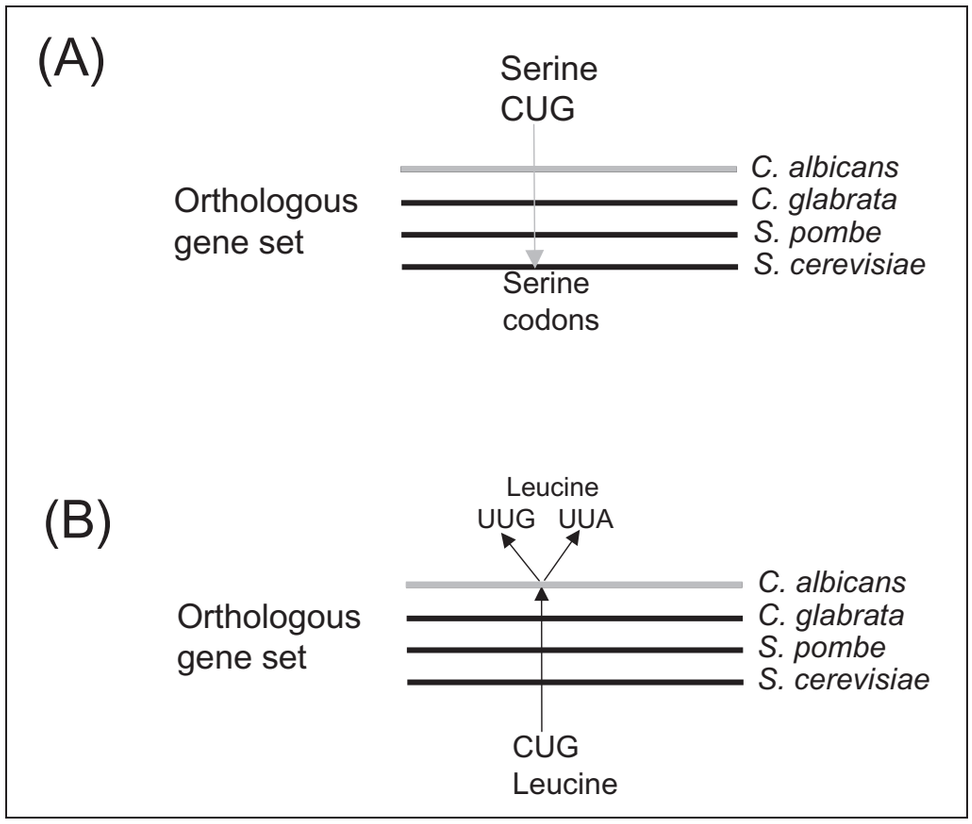


Figure 5: Comparative genomics permits tracing of the fate of codons undergoing identity change. Evolution of genetic code alterations through ambiguous codon decoding or biased GC pressure (codon capture) has a strong impact on the usage of codons that undergo identity change and also on codons belonging to its codon family. This effect was unveiled using comparative genomics methodologies.[25] For this, a set of orthologuous genes from *C. albicans*, *S. cerevisiae* and *S. pombe* were aligned and the level of conservation of the CUG codon in the three genomes was determined. The data showed that ambiguous CUG decoding drove most of the 'old' CUG codons from the genome of the *Candida* ancestor (A,B). This is clearly supported by the observation that CUG codons present in *C. albicans* genes are represented in the other genomes by serine and not leucine codons (A). The reciprocal alignment showed that CUG codons present in *S. cerevisiae* genes are represented by other leucine — but not serine — codons (mainly the leu–UUG codon) in the *C. albicans* homologous genes (B). Therefore, the 17,000 CUG codons existent in the genome of *C. albicans* are unrelated to the CUG codons existent in *S. cerevisiae* and *S. pombe* genomes and evolved during the period of redefinition of the identity of the CUG codon, ie during the last 272 million years[25]

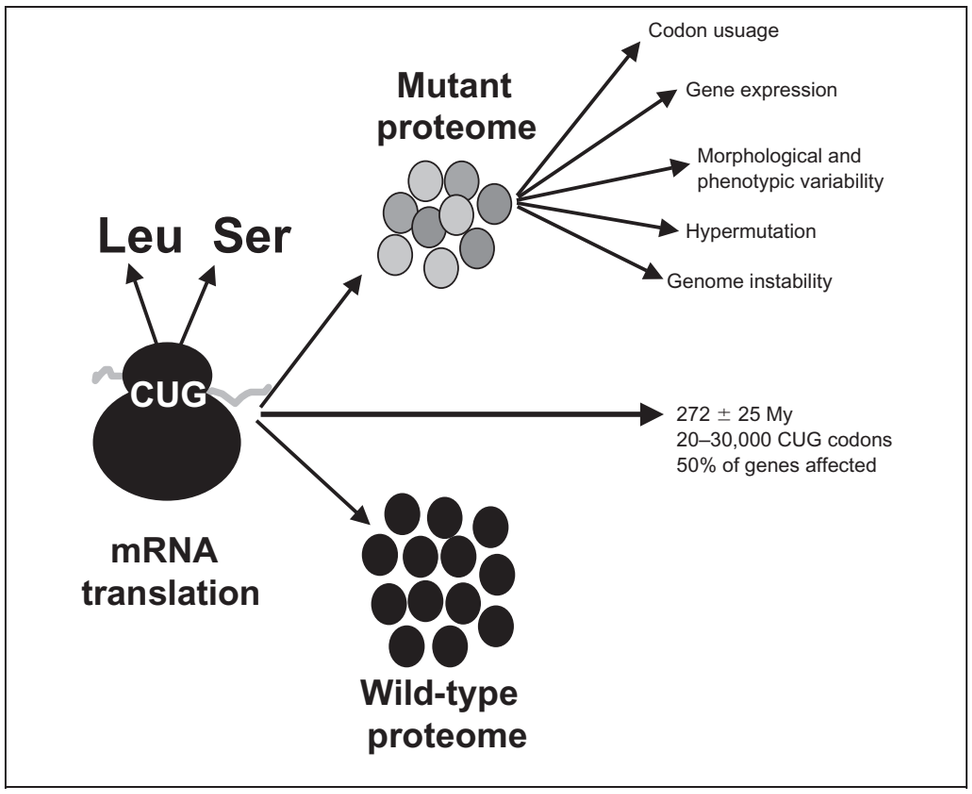


Figure 6: The global impact of genetic code ambiguity and codon identity redefinition. Codon identity redefinition through ambiguous decoding has a global impact on protein synthesis, resulting in proteome destabilisation. For example, the *C. albicans* 17,000 CUG codons are present in approximately 50 per cent of the genes at a frequency of one to ten codons per gene. In this case, such CUG ambiguity resulted in synthesis of a significant amount of mutant protein (mutant proteome) over approximately 272 million years (My). These mutant proteins were degraded or aggregated. Reconstruction of this genetic code change in *S. cerevisiae* showed that a significant portion of the mutant proteome is folded and remains active, although catalytic activity and stability is negatively affected. Additionally, large-scale protein misfolding/aggregation and degradation triggers the stress response — in particular, the expression of molecular chaperones causes morphological variation, induces a hypermutation condition and decreases genome stability[15,27,28]

Genetic code ambiguity has major physiological and evolutionary implications

The impact of ambiguous decoding on the usage of the CUG codon in *C. albicans* was elucidated by aligning a set of orthologous genes from *C. albicans*, *S. cerevisiae* and *S. pombe* and identifying the positions occupied by CUG codons in the alignment (Figure 5). This study[24] showed that CUG double identity (leucine and serine) forced the disappearance of approximately 30,000 CUG codons existent in the *Candida* ancestor; that is, like biased genome GC pressure, codon decoding ambiguity is able to drive codons from the genome.[25]

mRNA decoding ambiguity forces codons from the genome

Interestingly, a CUG scan of the *C. albicans* genome identified 17,000 CUG codons. This apparent contradiction between CUG disappearance in the ancestor and the existence of 17,000 CUG codons in the genome of *C. albicans* was resolved by reciprocal scoring of CUG codons in the alignment of the orthologous gene set (Figure 5A,B). The refined sequence alignment showed that only 2 per cent of all CUG codons present in the *C. albicans* genome are represented in the other genomes by one of the six leucine codons. The remaining

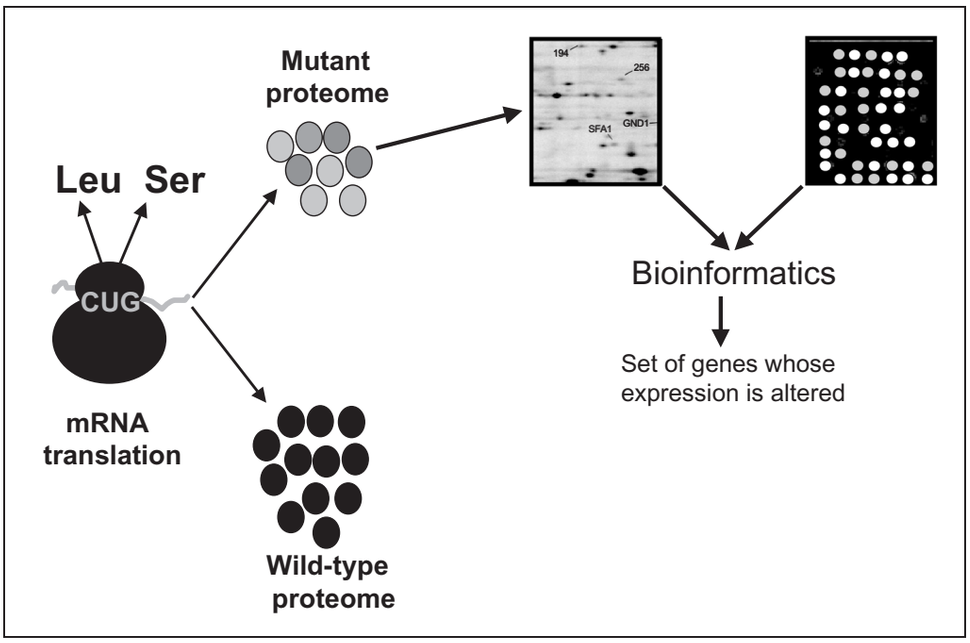


Figure 7: Functional genomics methodologies are powerful tools for studying the impact of codon identity redefinition and genetic code ambiguity on gene expression. Genetic code ambiguity in general, and codon identity redefinition in particular, are global phenomena whose impact can only be tackled using global gene expression methodologies — namely, proteomics, DNA microarrays and metabolomics. Large-scale synthesis of mutant proteins results in accumulation of misfolded and aggregated proteins and protein degradation. Cells respond to this by overshooting expression of molecular chaperones and the protein degradation machinery, ie the proteasome in eukaryotic cells. Accumulation of misfolded proteins — induced by ambiguous mRNA decoding — mimics stress conditions, namely heat shock,[29] triggering the stress response. The availability of these DNA microarrays and annotated two-dimensional maps for *S. cerevisiae* makes it an ideal model organism for studying both the effect of codon identity redefinition and genetic code ambiguity in general

Genetic code changes are better studied using global genomics methodologies

The *S. cerevisiae* and *C. albicans* CUG codons are not related

98 per cent of CUG codons appear in positions represented by serine codons or codons coding for conserved amino acids of serine (Figure 5A,B). That is, the 30,000 and 26,000 CUG codons present in the *S. cerevisiae* and *S. pombe* genomes, respectively, are not related to the 17,000 CUG codons present in the *C. albicans* genome, indicating that the latter codons — which are decoded as serine and not leucine — evolved during the last 272 My from codons coding for serine rather than leucine. This implies that CUG identity redefinition was mediated through a dynamic molecular mechanism that involved gradual disappearance of 'old' ambiguous leu–CUG codons (coding for leucine plus serine) and simultaneous emergence of 'new' ambiguous ser–CUG codons arising via mutation of codons that coded for serine or conserved amino acids of serine.[25]

This novel pathway of codon identity redefinition exposes in a rather dramatic manner an unsuspected flexibility in the genetic code and shows that *C. albicans*, and most likely all other organisms, are well equipped to tolerate the negative impact of proteome disruption caused by codon identity redefinition. Additionally, it implies that the *C. albicans* proteome has been unstable during the last 272±25 My because both the disappearing (old) and emerging (new) codons remained ambiguous since the appearance of the ser–$tRNA_{CAG}$ to the present day.[23] Interestingly, the appearance of the novel Ser–$tRNA_{CAG}$ was a major evolutionary

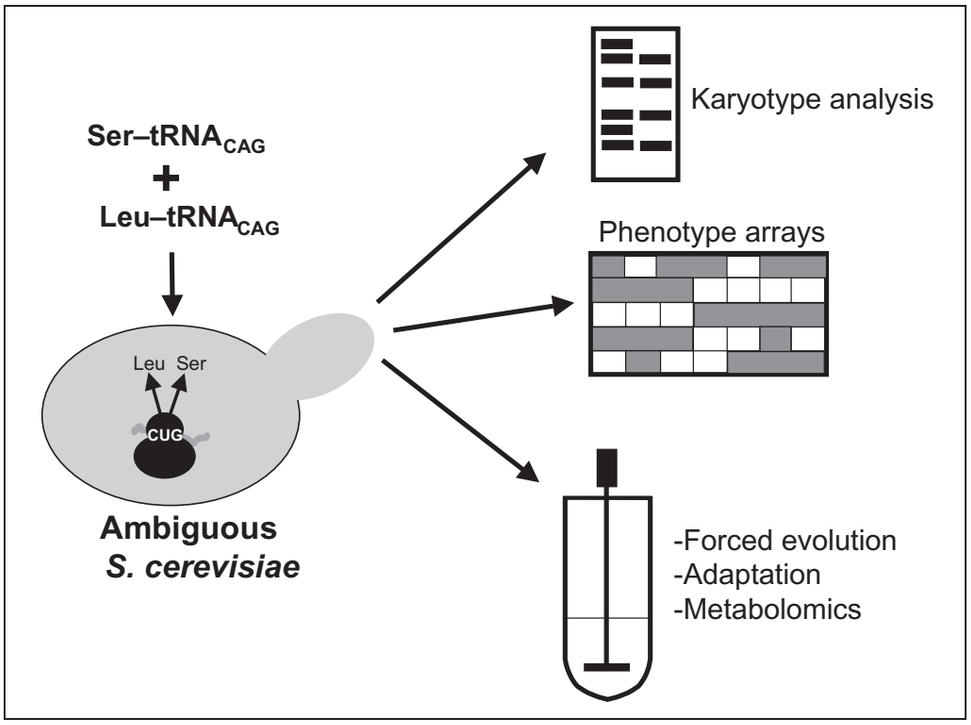


Figure 8: *S. cerevisiae* as a model for elucidating the physiological and evolutionary significance of genetic code alterations. The evolutionary significance of non-standard genetic codes can be elucidated by studying the impact of codon identity redefinition at physiological and ecological levels. The reconstruction of the evolutionary pathway of *Candida* CUG identity redefinition in *S. cerevisiae* is an excellent experimental model for tackling the high complexity associated with the global effects of ambiguous mRNA decoding. By studying ambiguous yeast cell lines using forced evolution methodologies and phenotype arrays, and by analysing their metabolome and karyotype, one is bound to provide new insights into the global effects of codon decoding ambiguity and identity redefinition on genome stability and physiology and, ultimately, in understanding at the molecular level the surprising selective advantage displayed by ambiguous cell lines in adaptation to environmental stress conditions and ultimately new ecological niches[15,28]

CUG decoding ambiguity is 272 ± 25 My old

force shaping the usage of all six leucine codons, ie UUA, UUG, CUA, CUG, CUC and CUU, thus showing that redefinition of the identity of a single codon affects the usage of all codons belonging to that codon family.[25]

FUNCTIONAL GENOMICS OF NON-STANDARD GENETIC CODES

Other important biological questions arising from the evolution of non-standard genetic codes through ambiguous codon decoding relate to the impact of proteome destabilisation on physiology, genetic variability and adaptation (Figure 6). The most relevant questions are:

- How do organisms cope with large-scale synthesis of mutant/aberrant proteins?
- What is the destiny of these proteins?
- What is the relevance of proteome destabilisation on physiology, gene expression, genome stability and, more importantly, the evolution of the species that define codon identity?

Reconstruction of the *Candida* genetic code change in *S. cerevisiae* is proving to be an excellent experimental approach for

The proteome of ambiguous cells is unstable

tackling these questions because of the availability of powerful genetics and genomics tools for *S. cerevisiae*. Additionally, the CUG codon is used at low frequency in both yeasts — there are 30,000 and 17,000 CUG codons in *S. cerevisiae* and *C. albicans*, respectively — and the *C. albicans* ser–$tRNA_{CAG}$ is correctly processed and charged in *S. cerevisiae*.[22,25] Such reconstruction, involving transformation of *S. cerevisiae* with the *C. albicans* ser–$tRNA_{CAG}$, creates ambiguous cell lines — the CUG codon is decoded as both leucine by the endogenous leu–$tRNA_{UAG}$ and serine by the transformed *C. albicans* ser–$tRNA_{CAG}$ — that can be used to study the impact of codon identity redefinition and ambiguity on physiology and adaptation using global genomics approaches, namely phenotypic arrays, DNA microarrays and proteomics (Figures 7 and 8).

Proteome destabilisation triggers the stress response

Preliminary data validate this experimental approach. For example, phenotypic characterisation of ambiguous *S. cerevisiae* cells showed that, despite having a slower growth rate, they are tolerant to heavy metals, drugs, ethanol, oxidants and sodium chloride.[15,29] In other words, despite their growth disadvantage observed in rich media, ambiguous cells undergoing codon identity redefinition have a selective advantage under specific environmental conditions — namely, extreme environmental stress conditions.[28] This surprising result suggests that accumulation of aberrant/unfolded proteins synthesised through ambiguous mRNA decoding triggers expression of stress proteins — namely, the molecular chaperones Hsp104 and Hsp70 — that protect ambiguous cells on exposure to severe stress.[28]

Since ambiguous mRNA decoding destabilises a significant part of the proteome, which results in permanent overexpression of stress proteins (proteins normally expressed under stress conditions or in stationary phase), this permanent alteration in gene expression results in pleiotropic phenotypic effects. Therefore, the redefinition of the CUG identity over 272 My is likely to have had a major impact on the evolution of the genus *Candida*. The same principle should apply to all organisms that use non-standard genetic codes. Ongoing studies in the authors' laboratory aim at elucidating these questions.

CONCLUSIONS

Despite the discovery of the first genetic code change more than 30 years ago, it is only now that researchers are beginning to understand the functional and evolutionary significance of non-standard genetic codes. Molecular phylogeny data clearly show that non-standard genetic codes evolve from the standard genetic code,[1–5] excluding the hypothesis that they represent molecular relics of alternative codes that may have existed in the RNA world. The full evolutionary implications of changing the genetic code are far from being understood and one is urged to develop methodologies, such as those outlined in this paper, to tackle their study experimentally.

The full diversity of the genetic code is also far from being understood; however, the diversity already uncovered suggests that it is likely that new alterations to the standard code will be discovered in the future. This has practical implications for genome annotation, gene homology studies and heterologous gene expression. For example, non-standard genetic codes introduce significant noise in protein sequence alignments and, more importantly, genes from organisms with non-standard genetic codes may not express functional proteins in host organisms with the standard genetic code. This is well illustrated by the expression of certain *C. albicans* genes in *S. cerevisiae* and also by the converse expression of heterologous reporter genes in *C. albicans*, namely the commonly used green fluorescent protein (GFP), which is non-functional in *C. albicans*.[29]

Finally, reconstruction of the *Candida* genetic code change in its close relative

Reconstruction of the *C. albicans* genetic code change in *S. cerevisiae* provides a unique model to shed new light on the evolution of codon identity redefinition

S. cerevisiae is proving to be a very powerful tool in dissecting the molecular mechanisms and elucidating the functional implications of this genetic code change for the evolution of the genus *Candida*. This, and expansion of the genetic code through genetic engineering approaches, clearly shows that the genetic code is flexible and that organisms are well equipped to deal with genetic code ambiguity.[30–34] More importantly, these pioneering experiments define an experimental framework for shedding new light on, and hopefully elucidating, how and why certain organisms evolve alternative genetic codes.

References

1. Barrell, B. G., Bankier, A. T. and Drouin, J. (1979), 'A different genetic code in human mitochondria', *Nature*, Vol. 282, pp. 189–194.
2. Horowitz, S. and Gorovsky, M. A. (1985), 'An unusual genetic code in nuclear genes of *Tetrahymena*', *Proc. Natl. Acad. Sci. USA*, Vol. 82, pp. 2452–2455.
3. Meyer, F. M., Schmidt, H. J., Plumper, E. *et al.* (1991), 'UGA is translated as cysteine in pheromone 3 of *Euplotes octocarinatus*', *Proc. Natl. Acad. Sci. USA*, Vol. 88, pp. 3758–3761.
4. Yokobori, S., Suzuki, T. and Watanabe, K. (2001), 'Genetic code variations in mitochondria: tRNA as a major determinant of genetic code plasticity', *J. Mol. Evol.*, Vol. 53, pp. 314–326.
5. Knight, R. D., Freeland, S. J. and Landweber, L. F. (2001), 'Rewiring the keyboard: evolvability of the genetic code', *Nat. Rev. Genetics*, Vol. 2, pp. 49–58.
6. Osawa, S. and Jukes, T. H. (1989), 'Codon reassignment (codon capture) in evolution', *J. Mol. Evol.*, Vol. 28, pp. 271–278.
7. Osawa, S., Jukes, T. H., Watanabe, K. and Muto, A. (1992), 'Recent evidence for the evolution of the genetic code', *Microbiol. Rev.*, Vol. 56, pp. 229–264.
8. Osawa, S. (1995), 'Evolution of the genetic code', Oxford University Press, Oxford, UK.
9. Schultz, D. W. and Yarus, M. (1994), 'Transfer RNA mutation and the malleability of the genetic code', *J. Mol. Biol.*, Vol. 235, pp. 1377–1380.
10. Schultz, D. W. and Yarus, M. (1994), 'tRNA structure and ribosomal function. I. tRNA nucleotide 27–43 mutations enhance first position wobble', *J. Mol. Biol.*, Vol. 235, pp. 1381–1394.
11. Schultz, D. W. and Yarus, M. (1996), 'On the malleability in the genetic code', *J. Mol. Evol.*, Vol. 42, pp. 597–601.
12. Kano, A., Andachi, Y., Ohama, T. and Osawa, S. (1991), 'Novel anticodon composition of transfer RNAs in *Micrococcus luteus*, a bacterium with a high genomic G + C content: correlation with codon usage', *J. Mol. Biol.*, Vol. 221, pp. 387–401.
13. Goffeau, A., Smith, B. T., Lewis, G. *et al.* (1997), 'The yeast genome directory', *Nature*, Vol. 387(suppl), pp. 5–105.
14. Knight, R. D., Landweber, L. F. and Yarus, M. (2001), 'How mitochondria redefine the code', *J. Mol. Evol.*, Vol. 53, pp. 299–313.
15. Santos, M. A. S., Ueda, T. and Tuite, M. F. (1997), 'The non-standard genetic code of *Candida* spp: an evolving genetic code switch or a novel mechanism for adaptation?' *Mol. Microbiol.*, Vol. 26, pp. 423–431.
16. Lehman, N. (2001), 'Molecular evolution: please release me, genetic code', *Curr. Biol.*, Vol. 11, pp. R63–R66.
17. Murgola, E. (1985), 'tRNA suppression, and the code', *Ann. Rev. Genet.*, Vol. 19, pp. 57–80.
18. Santos, M. A. S., Cheesman, C., Costa, V. *et al.* (1999), 'Selective advantages created by codon ambiguity allowed for evolution of an alternative genetic code in *Candida* spp', *Mol. Microbiol.*, Vol. 31, pp. 937–947.
19. Kawaguchi, Y., Honda, H., Taniguchi-Morimura, J. and Iwasaki, S. (1989), 'The codon CUG is read as serine in an asporogenic yeast *Candida cylindracea*', *Nature*, Vol. 341, pp. 164–166.
20. Santos, M. A. S. and Tuite, M. F. (1995), 'The CUG codon is decoded *in vivo* as serine and not leucine in *Candida albicans*', *Nucleic Acids Res.*, Vol. 23, pp. 1481–1486.
21. Ohama, T., Suzuki, T., Mori, M. *et al.* (1993), 'Non-universal decoding of the leucine codon CTG in several *Candida* species', *Nucleic Acids Res.*, Vol. 21, pp. 4039–4045.
22. Santos, M. A. S., Perreau, V. and Tuite, M. F. (1996), 'Transfer RNA structural change is a key element in the reassignment of the CUG codon in *Candida albicans*', *EMBO J.*, Vol. 15, pp. 5060–5068.
23. Suzuki, T., Ueda, T. and Watanabe, K. (1997), 'The 'polysemous' codon – a codon with multiple amino acid assignment caused by dual specificity of tRNA identity', *EMBO J.*, Vol. 16, pp. 1122–1134.
24. Tuite, M. F. and Santos, M. A. S. (1997), 'Codon reassignment in *Candida* species: an evolutionary conundrum', *Biochimie*, Vol. 78, pp. 993–999.
25. Massey, S. E., Moura, G., Beltrão, P. *et al.* (2003), 'Comparative evolutionary genomics

unveils the molecular mechanisms of reassignment of the CTG codon in *Candida* spp', *Genome Res.*, Vol. 13, pp. 544–557.

26. Pesole, G., Lotti, M., Alberghina, L. and Saccone, C. (1995), 'Evolutionary origin of non-universal CUG^{Ser} codon in some *Candida* species as inferred from molecular phylogeny', *Genetics*, Vol. 141, pp. 903–907.
27. Grant, C. M, Firoozan, M. and Tuite, M. F. (1989), 'Mistranslation induces the heat-shock response in the yeast *Saccharomyces cerevisiae*,' *Mol. Microbiol.*, Vol. 3, pp. 215–220.
28. Siderius, M. and Mager, W. H. (1997), 'General stress response: in search of a common denominator', in Hohmann, S. and Mager, W. H., Eds., 'Yeast Stress Responses', Chapman and Hall, New York, pp. 213–224.
29. Cormack, B. P., Bertram, G., Egerton, M. *et al.* (1997), 'Yeast-enhanced green fluorescent protein (yEGFP): a reporter of gene expression in *Candida albicans*', *Microbiology*, Vol. 143, pp. 303–311.
30. Schimmel, P. and Marlière, P. (2001), 'Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway', *Science*, Vol. 292, pp. 501–504.
31. Wang, L., Brock, A., Herberich, B. and Schultz, P. G. (2001), 'Expanding the genetic code of *Escherichia coli*', *Science*, Vol. 292, pp. 498–500.
32. Wang, L., Zhang, Z., Brock, A. and Schultz, P. G. (2003), 'Addition of the keto functional group to the genetic code of *Escherichia coli*', *Proc. Natl. Acad. Sci. USA*, Vol. 100, pp. 56–61.
33. Doring, V., Mootz, H. D., Nangle, L. A. *et al.* (2001), 'Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway', *Science*, Vol. 292, pp. 501–504.
34. Sacks, M. E. (2001), 'Making sense out of nonsense', *Proc. Natl. Acad. Sci. USA*, Vol. 98, pp. 2125–2127.