Universidade de Aveiro Departamento de Biologia
2013

Ana Rita Macedo Bezerra

Genómica molecular de uma alteração ao código genético.

Molecular genomics of a genetic code alteration.

Universidade de Aveiro Departamento de Biologia
2013

Ana Rita Macedo Bezerra

Genómica molecular de uma alteração ao código genético.

Molecular genomics of a genetic code alteration.

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Biologia, realizada sob a orientação científica do Prof. Doutor Manuel António da Silva Santos, Professor Associado do Departamento de Biologia da Universidade de Aveiro

Apoio financeiro da FCT e do FSE no âmbito do III Quadro Comunitário de Apoio.

o júri

presidente

Doutora Maria Hermínia Deulonder Correia Amado Laurel
Professora Catedrática da Universidade de Aveiro

Doutora Judith Berman
Professora Catedrática da Universidade de Tel Aviv

Doutora Margarida Paula Pedra Amorim Casal
Professora Catedrática da Universidade do Minho

Doutor Manuel António da Silva Santos
Professor Associado da Universidade de Aveiro

Doutora Isabel Antunes Mendes Gordo
Investigadora Principal do Instituto Gulbenkian de Ciência

Doutor António Carlos Matias Correia
Professor Catedrático da Universidade de Aveiro

agradecimentos

acknowledgements

First and foremost, I would like to thank my supervisor, Doutor Manuel Santos, for the opportunity to work on this project and for his support throughout the last 5 years. Thank you for keeping me going when times were tough, asking insightful questions, and offering invaluable advice whilst allowing me the room to work in my own way.

I am indebted to many colleagues who helped me during these last 5 years, especially to João Simões whose precious input was essential for this work. To Tatiana, Cristina, Denisa and João Paredes for helping me when I started my laboratory work. To Ana, Marisa, Tobias, Laura, Céu and Patrícia for support and fruitful discussions during my journey in the laboratory. To Violeta for the constant patience and readiness with everything and to the other past members of the Aveiro RNA Biology laboratory for their friendship and for making the laboratory such a pleasant place to do science.

My gratitude goes also to my international colleagues of the EU-FP7 Sybaris consortium that helped with the *Candida albicans* whole genome sequencing and data analysis. In particular, Wanseon Lee and Johan Rung from the European Bioinformatics Institute (EBI-EMBL) in Cambridge, Ivo Gut and Mónica Bayés from Centro Nacional de Análises Genómico (CNAG) in Barcelona and Lisa Rizzetto from Florence University in Florence. And also to Professor Judith Berman from the Universities of Telavive and Minnesota for the constructive comments and discussions about my work.

I am grateful to FCT, for funding my work through a PhD grant SFRH/BD/39030/2007 and project PTDC/SAU-GMG/098850/2008.

Finally, I am deeply grateful to my parents, who stood by me and supported my choices and to my sister for her encouragement and care. A very special thanks to all my friends, who were always a phone call away, in particular to Renato for all the listening and advising.

A big thank you to all of you.

palavras-chave

Candida albicans, código genético, tRNA, evolução, erros de tradução, transcriptoma, genoma

resumo

O código genético não é universal. Alterações à identidade de vários codões descobertas em procariotas e eucariotas invalidam a hipótese dum código genético universal e imutável. Por exemplo, várias espécies do género *Candida* traduzem o codão CUG de leucina como serina. Em *Candida albicans*, um único tRNA de serina ($tRNA_{CAG}^{Ser}$) incorpora *in vivo* 97% de serina e 3% de leucina nas proteínas em resposta a codões CUG presentes nos mRNAs deste fungo patogénico. Esta ambiguidade é flexível e a incorporação de leucina aumenta em condições de stress. De forma a elucidar a função desta ambiguidade e determinar se a identidade dos codões CUG podia ser revertida de serina para leucina, desenvolvemos uma estratégia de evolução forçada e uma proteína recombinante fluorescente cuja actividade depende da incorporação de leucina num codão CUG. Construímos estirpes que incorporam leucina nas proteínas em resposta a codões CUGs em níveis que variam entre 0,64% e 98,46%. Esta reversão de uma alteração ao código genético demostrou de modo inequívoco que o código é flexível e pode evoluir. Testes de crescimento em diferentes meios de cultivo revelaram uma série impressionante de fenótipos com elevado potencial adaptativo nas estirpes mais ambíguas, nomeadamente tolerância a antifúngicos. A sequenciação dos genomas das estirpes que construímos revelou que a ambiguidade do codão CUG resulta na acumulação de polimorfismos de nucleótido únicos (SNP) no genoma. Verificámos também perda de heterozigozidade (LOH) nos cromossomas 5 e R das estirpes que incorporam 80,84% e 98,46% de leucina em locais proteicos codificados por codões CUG. Os SNPs acumularam-se preferencialmente em genes envolvidos na adesão celular, no crescimento filamentoso e na formação de biofilmes, sugerindo que *C. albicans* utiliza a sua ambiguidade natural para aumentar a diversidade genética dos processos relacionados com a patogénese e resistência a drogas. Estes resultados evidenciam uma notável flexibilidade do código genético de *C. albicans* e revelam funções inesperadas da ambiguidade do código genético na evolução da diversidade genética e fenotípica.

keywords

Candida albicans, genetic code, tRNA, evolution, mistranslation, transcriptome, genome

abstract

The genetic code is not universal. Alterations to its standard form have been discovered in both prokaryotes and eukaryotes and demolished the dogma of an immutable code. For instance, several *Candida* species translate the standard leucine CUG codon as serine. In the case of the human pathogen *Candida albicans*, a serine tRNA ($tRNA_{CAG}^{Ser}$) incorporates *in vivo* 97% of serine and 3% of leucine in proteins at CUG sites. Such ambiguity is flexible and the level of leucine incorporation increases significantly in response to environmental stress. To elucidate the function of such ambiguity and clarify whether the identity of the CUG codon could be reverted from serine back to leucine, we have developed a forced evolution strategy to increase leucine incorporation at CUGs and a fluorescent reporter system to monitor such incorporation *in vivo*. Leucine misincorporation increased from 3% up to nearly 100%, reverting CUG identity from serine back to leucine. Growth assays showed that increasing leucine incorporation produced impressive arrays of phenotypes of high adaptive potential. In particular, strains with high levels of leucine misincorporation exhibited novel phenotypes and high level of tolerance to antifungals. Whole genome re-sequencing revealed that increasing levels of leucine incorporation were associated with accumulation of single nucleotide polymorphisms (SNPs) and loss of heterozygozity (LOH) in the higher misincorporating strains. SNPs accumulated preferentially in genes involved in cell adhesion, filamentous growth and biofilm formation, indicating that *C. albicans* uses its natural CUG ambiguity to increase genetic diversity in pathogenesis and drug resistance related processes. The overall data provided evidence for unantecipated flexibility of the *C. albicans* genetic code and highlighted new roles of codon ambiguity on the evolution of genetic and phenotypic diversity.

List of contents

List of figures

List of tables

List of abbreviations

AARS	aminoacyl-tRNA synthetase
aa-tRNA	aminoacyl-tRNA
APS	ammonium persulphate
ARG4	arginine biosynthesis gene
ATP	adenosine 5'-triphosphate
cDNA	cyclic deoxyribonucleic acid
CFU	colony forming units
$CuSO_4$	copper sulfate
DEAE	diethylaminoethyl
DNA	deoxyribonucleic acid
DNase	deoxyribonulease
dNTP	deoxynucleotide solution mix
DTT	dithiothreitol
EDTA	ethylenediamine tetraacetic acid
ESR	environmental stress response
g (mg, µg, ng)	gram (milligram, microgram, nanogram)
GO	gene ontology
HIS1	histidine biosynthesis gene
HSP	heat-shock protein
kDa	kilodalton
L (mL, µL)	liter (mililiter, microliter)
LiCl	lithium chloride
LOH	loss of heterozygosity
M (mM, µM)	molar (milimolar, micromolar)
mA	miliampère
mQ	milliQ
mRNA	messenger ribonucleic acid
NaCl	sodium chloride
NADH	nicotinamide adenine dinucleotide
NaH_2PO_4	sodium dihydrogen phosphate

nt	nucleotide
OD_{600}	optical density at 600nm
ORF	open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PM	phenotype microarray
PMSF	phenylmethanesulfonylfluoride
RF1	release factor-1
RNA	ribonucleic acid
ROS	reactive oxygen species
rpm	revolutions per minute
RPS10	ribosomal protein S10
rRNA	ribosomal ribonucleic acid
RT	reverse transcription
SAPK	stress-activated protein kinase pathway
SDS	sodium dodecyl sulphate
SNP	single nucleotide polymorphism
SPI	selective pressure incorporation
TEMED	tetramethylethylenediamine
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid
UTR	untranslated region
UV	ultraviolet
V	volt
*yE*GFP	yeast-enhanced green fluorescent protein
YEPD	yeast extract peptone dextrose

Other abbreviations will be explained when used in the text.

1. Introduction

1.1 Genetic code: the interface between nucleic acid and protein chemistry

1.1.1 The standard genetic code

All organisms are shaped by a genetic program that determines all their biological activities. This program operates in a way that the information flow between nucleic acids and protein is unidirectional and is accomplished via two processes: transcription and translation. During transcription the information inherent to DNA is copied to mRNA molecules that are finally translated into an amino acid sequence by the ribosome. The connection between these processes was incorporated very early in the concept termed “central dogma of molecular biology” (Crick, 1958). In order to bring these two realms together, the genetic code establishes a set of rules that relates precisely the sequence of base triplets in nucleic acid with a sequence of amino acid residues in protein (Crick, 1958;Nirenberg and Leder, 1964).

The genetic code arrangement as presented in Figure 1-1 is known as the “universal code” and has several basic characteristics. First, it maps 64 codons onto a set of 20 amino acids and the stop signal. To do so, each codon is a triplet of bases – two pyrimidines (Uridine, Cytosine) and two purines (Adenine, Guanine) - thus forming the 64 (4^3) codon set (Nirenberg *et al.*, 1966;Crick, 1968). While methionine and tryptophan are assigned only with one coding triplet (AUG and UGG, respectively), most of the amino acids are represented by more than one codon. Therefore, the standard code is highly redundant in its structure. The chemical diversity encoded in this way includes carboxyl groups (Glu, Asp), amides (Gln, Asn), a thiol (Cys) and a thiol ester (Met), alcohols (Ser, Thr), a basic amine (Lys), guanidine (Arg), aliphatic (Ala, Val, Leu, Ile) and aromatic moieties (Phe, Tyr, His, Trp), and also a cyclic imino acid residue (Pro) (Ardell and Sella, 2001).

Furthermore, the codons for a single amino acid are arranged in blocks, rather than being randomly distributed throughout the code (with the exception of serine). Amino acids with two-codon or four-codon boxes differ only at the third position of the codon. In cases of six codons (arginine and leucine), these form one four-codon box and one two-codon box that differ in the first position (Figure

1-1). The importance of the nucleotides in the three codon positions significantly diverges: 69% of the mutations in the third codon are synonymous, only 4% of the mutations in the first position are synonymous, and none of the mutations in the second position are synonymous (Taylor and Coates, 1989). Another inherent part of the genetic code is the dual partitioning of polar and non polar amino acids. All codons with a U in the second position are cognate to amino acids with hydrophobic side-chains and codons with a central A are cognate to amino acids with chemically variable, polar side-chains (Figure 1-1). This fact is of critical importance, since hydrophobicity is an essential force in protein folding (Ardell and Sella, 2002).

Generally, the structure of the code appears to be robust enough to conform with the principles of optimal coding (Novozhilov *et al.*, 2007).

	U		C		A		G	
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
	UUA	Leu	UCA	Ser	UAA	TER	UGA	TER
	UUG	Leu	UCG	Ser	UAG	TER	UGG	Trp
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Figure 1-1. The "universal" genetic code. The codon series are shaded in accordance with different characteristics of amino acids. Saturation reflects molecular volume; colourful equals bigger. Brightness reflects polar requirement; lighter equals hydrophobic. Hue reflects side-chain composition; 0º (red) = acid; 30º (orange) = amide; 60º (yellow) = sulphur; 120º (green) = alcohol; 180º (cyan) = aromatic; 240º (blue) = basic; 270º (purple) = hydrophobic. Adapted from Knight *et al.*, 1999.

1.1.2 Genetic code components: AARS and tRNAs

The precision of the translational process depends for the most part on the accuracy of aminoacylation (covalent connection of cognate amino acids with cognate transfer RNAs) and consequent interactions between charged tRNA with ribosome-bound mRNA (Ibba and Soll, 1999). Indeed, it is the specific attachment of amino acids onto tRNA adaptor molecules by amynoacyl-tRNA synthetases (AARSs) that defines the genetic code (Woese *et al.*, 2000;Giege, 2006;Giege, 2008). This suggests that the reconstruction of the evolutionary pathways that established the genetic code requires deep knowledge of AARSs, tRNAs, mRNAs and the ribosome.

1.1.2.1 Molecular architecture of AARS

AARSs are the real translators of the genetic code because they are responsible for the specific esterification of tRNAs with their cognate amino acids. Several investigations have shown that prokaryotic AARS appear as free cytosolic enzymes rarely bound with other proteins. However, the eukaryotic AARS enzymes are usually larger due to the presence of the C- and N- terminal extensions that are unnecessary for aminoacylation (Doolittle and Handy, 1998) and recent studies show that eukaryotic AARS are involved in a wide repertoire of functions: translational fidelity, tRNA processing, RNA splicing, RNA trafficking, apoptosis and transcriptional and translational regulation (Martinis *et al.*, 1999a;Martinis *et al.*, 1999b).

AARSs constitute a family of 20 cellular enzymes that are believed to have been present in the last universal common ancestor of the tree of life (Schimmel, 2008b). During the evolutionary process, AARSs were classified into two different structural families (class I and class II) based on their different size, sequence analyses and on their three-dimensional structures in free and complex forms (Eriani *et al.*, 1990). Of the twenty AARSs, 10 are found in each family (Figure 1-2) and they differ as much as they reveal structurally distinct topologies within the

active site (domain that recognizes the acceptor stem of the tRNA, where the amino acid is attached) (Schimmel and Ribas de, 1995). All the enzymes in each class share a distinctive active domain, with the exception of LysRS that exists as a class I or as a class II enzyme in different organisms (Ibba *et al.*, 1997b).

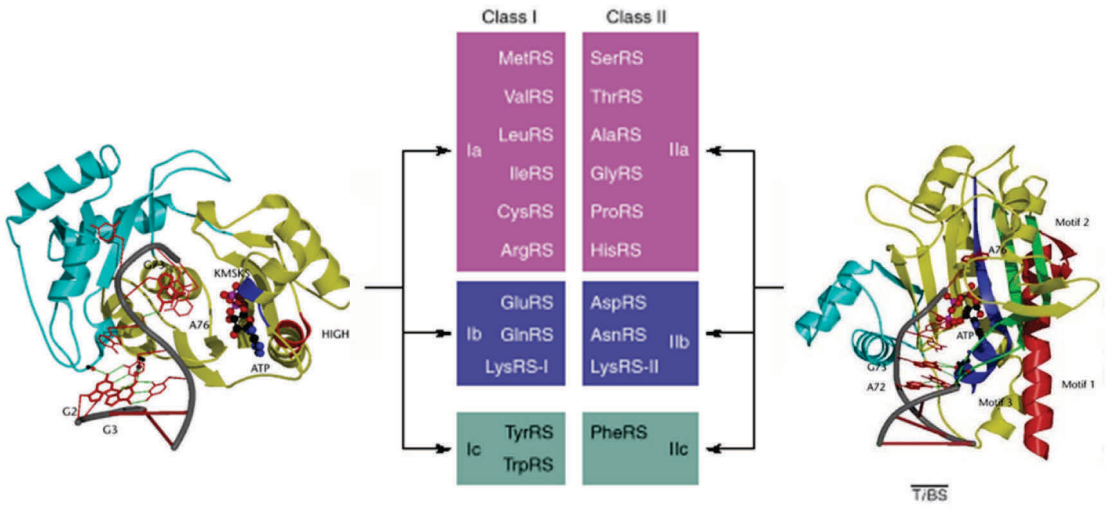


Figure 1-2. Classes of aminoacyl tRNA synthetases. The two major families can be organized into subclasses that enclose enzymes that are closely related to each other in their sequences. Notably, the subclasses group AARSs according to their amino acid chemical types (Schimmel, 2008b).

All members of class I AARSs share an active-site domain that forms a Rossmann nucleotide-binding fold (formed by a five-handed parallel β-sheet), already found in NADH-binding dehydrogenases and ATP-binding kinases (Giege, 2006). Also, in class I enzymes there are two conserved sequences: HIGH and KMSKS. These sequence motifs are vital elements in the structure of the active site for aminoacyl-adenylate synthesis (Arnez and Moras, 1997;Cusack, 1997). Confirming this, several mutagenesis studies of the KMSKS sequence in TyrRS, MetRS and TrpRS revealed that these motifs are part of a mobile loop that stabilizes the transition state of the activation reaction. Similarly, the same role was attributed for the HIGH sequence in MetRS and TrpRS. In these cases, the amino acid and ATP binding and activation are accompanied by distinct

conformational alterations through the movement of small domains that hold these motifs (First and Fersht, 1993).

Class II enzymes have a unique topology, completely different from that of class I enzymes. Their structure was firstly identified in SerRS from *E. coli* (Cusack *et al.*, 1990) and the main feature is a seven-stranded β-sheet fold with three sequence motifs: motif 1 (18 amino acids), 2 (23–31 amino acids) and 3 (29–34 amino acids). Motif 1 holds an invariant Pro residue implicated in dimmer formation, while motives 2 and 3 are involved in aminoacyl-adenylate formation and binding of the 3'-end of tRNA (Arnez and Moras, 1997;Cusack, 1997).

Within each class, structural classifications organize AARSs into three distinct subclasses (Figure 1-2). Subclass Ia encloses enzymes that are specific for the amino acids leucine, isoleucine, valine, methionine, cysteine, and arginine; subclass Ib enzymes recognize glutamate, glutamine and lysine; subclass Ic are specific for tyrosine and tryptophan. Likewise, subclass IIa enzymes are specific for serine, threonine, glycine, alanine, proline, and histidine; subclass IIb recognize aspartate, asparagine, and lysine; subclass IIc holds the enzyme specific for phenylalanine (Cusack, 1997). Remarkably, subclasses c have synthetases for aromatic amino acids, subclasses b capture the carboxyl side-chain amino acids and the amidated (NH2) derivatives. Finally, subclass a has many of the hydrophobic amino acids (Figure 1-2).

These structural observations are consistent with early work showing that both classes also differ because class I AARSs approach the end of tRNA acceptor helix from the minor groove side and catalyse attachment of the amino acid to the 2'-OH at the end of tRNA chain. On the opposite, class II AARSs approach the end of tRNA from the major groove, attaching the amino acid to the 3'-OH. Significant exceptions are class Ic enzymes TyrRS and TrpRS that bind the tRNA from the major groove side and catalyze the attachment of the amino acid to the 2'-OH or 3'-OH of the tRNA. Also, class IIc enzyme PheRS binds the tRNA on the minor groove side and catalyzes the attachment of the amino acid to the 2'-OH (Sprinzl and Cramer, 1975).

A full set of AARS genes was found in all eukaryotes, but some archaea and bacteria do not have genes for all 20 AARS, nevertheless they use the

canonical 20 amino acids to construct their proteins. For example, AsnRS and GlnRS are missing in many taxa of the Archaea and bacteria and synthesis of Asn and Gln require the recruitment of enzymes of intermediary metabolism. Indeed, nondiscriminating aspartyl-tRNA synthetases (AspRSs) and glutamyl-tRNA synthetases (GluRSs) are required for Asn and Gln biosynthesis, respectively. The nondiscriminating AspRSs acylates both tRNA(Asp) and tRNA(Asn) with aspartate while the nondiscriminating GluRSs acylates both tRNA(Glu) and tRNA(Gln) with glutamate. Consequently, these enzymes provide misacylated aminoacyl-tRNA species (Asp-tRNA^{Asn} and Glu-tRNA^{Gln}) that are then modified into Asn-tRNA^{Asn} and Gln-tRNA^{Gln} through a phosphorylation/transamidation mechanism catalyzed by enzymes called tRNA-dependent amidotransferases (AdTs). Therefore, AdTs constitute essential enzymes for Asn-tRNA^{Asn} and Gln-tRNA^{Gln} synthesis in organisms lacking glutaminyl- and asparaginyl-tRNA synthetase (Feng *et al.*, 2005a;Feng *et al*., 2005b). In some other cases, the lack of an AARS is compensated by a dual function of another synthetase, as in the case of CysRS whose absence is compensated by the dual function of ProRS. In such organisms, biochemical and genetic analyses indicated that the prolyl-tRNA synthetase can indeed synthesize both Cys-tRNA^{Cys} and Pro-tRNA^{Pro} (Stathopoulos *et al.*, 2000). Therefore, the differential number of AARS in different organisms can be explained by the doubling of some AARS genes, postaminoacylation enzymatic modification of aminoacylated tRNA (AsnRS and GlnRS) or alternative decoding mechanisms as in Selenocysteine and Pyrrolysine (see 1.3.3).

1.1.2.2 Structure and function of the tRNA molecule

tRNA molecules link the RNA and the protein worlds, because they carry amino acids to the ribosomes where the triplet code of the mRNA is deciphered, originating the corresponding protein. Due to the redundancy of the genetic code, in the majority of organisms, there are more codons than tRNAs and more tRNAs than AARSs. Different tRNAs chargeable with the same amino acid are called

isoaccepting tRNAs and their number varies with the organism. For example, minimal genomes have only one tRNA for each amino acid, while eubacteria have 40-55 tRNA species (Márquez V. and Nierhaus K.H., 2004).

Early studies carried out by Robert Holley showed that tRNAs are small molecules (MW ~ 25-30 kDa) composed of a single polynucleotide chain with 72-93 ribonucleotides (Holley *et al.*, 1965). The exception is the nematode mitochondria that possess very short tRNAs (56 nucleotides) that lack either the D or the TψC stem loop (Watanabe *et al.*, 1994).

Comparison of 932 sequences demonstrated that despite considerable differences in primary structure, all tRNAs could assume the same "cloverleaf" secondary structure, as illustrated in Figure 1-3A (Sprinzl *et al.*, 1998). The two-dimensional model has four base-paired helices that define the four main regions of the molecule. At the top of the diagram, the acceptor stem is formed by seven base pairs linking the 5' and 3' -ends of the molecule. It has a conserved 5'-N^{73}CAA-3' sequence with a specific nucleotide at position 73 functioning as a discriminator base for amynoacylation specificity (see 1.1.2.4) (Giege *et al.*, 1998). The CCA end is the docking site of tRNAs at the A- and P-sites of the ribosomal peptidyl-transferase center (see 1.2.1). In most organisms (eukaryotes) the CCA end is added post-transcriptionally with a ATP (CTP): tRNA nucleotidyl-transferase or CCase (Augustin *et al*., 2003). In other cases (bacteria), the 3'-terminal CCA is encoded in the tRNA genes. The left domain of the cloverleaf structure contains the dihydrouracil loop (D loop), where the helical region has 3-4 base pairs and the loop has 8-11 nucleotides. The right side domain contains the T loop which is formed by 5 base pairs with a 7 nucleotides loop, characterized by the presence of a TψC sequence (T stands for ribose-thymidine and ψ for pseudouridine). At the opposite end of the acceptor stem is the anticodon loop with a universal length of 7 nucleotides and the anticodon in the middle. It also has a consensus sequence Py32 -U33 -XYZ-Pu (modified)-N38, where Py represents a pyrimidine, XYZ is the anticodon, Pu a purine base and N any nucleotide. The position 33 of the anticodon loop contains a conserved U in all tRNAs which is critical for the "U-turn" (a 180º turn which provides the correct geometry for tRNA interactions to the P and A sites in the ribosome (Márquez V. and Nierhaus K.H., 2004). Finally,

between the T-loop and the anticodon loop, there is a variable region, which can have 4-24 nucleotides. Based on the length of this region, tRNAs are classified as class I and class II. Class I includes the majority of tRNAs which contain 4-5 nucleotides. Class II tRNAs have longer variable arms of 10-24 nucleotides (eubacterial $tRNA^{Leu}$, $tRNA^{Ser}$ and $tRNA^{Tyr}$ and in some organelles) (Dirheimer *et al.*, 1995;Giege *et al.*, 1998).

All tRNAs have basically the same L-shaped tertiary structure, which results from stacking of the acceptor stem and the T stem loop, forming the short arm of the L, and the anticodon stem loop and the D stem loop forming the long arm (Figure 1-3 B and C). The extremities of both domains represent the functional domains of the tRNAs: the amino acid is attached to the tip of the acceptor stem, while the anticodon arm at the other extremity contains the anticodon triplet. This peculiar L-shape structure of tRNAs is stabilized largely by secondary interactions that form base-paired regions, but also by dozens of tertiary interactions between the T and D arms (base–base, base–backbone, and backbone–backbone interactions). The majority of the base-base tertiary interactions occur between invariant or semi-invariant bases, and permit the tRNA to fold and function correctly (Weaver, 2001).

However, tRNAs have almost 25% of modified nucleotides. The majority of those modifications are synthesized by specific enzymes during the maturation of tRNAs and include ribose/base methylations (Gm, Cm / m^5C), base isomerization (U to pseudouridine ψ), base reduction (U to D; dihydro-uridine), base thiolation (s^2C, s^2U, s^4U) and base deamination (inosine). The exception is queuosine at position 34 of tRNAs that read NAU or NAC codons (where N is any codon). In this case, modification involves an enzyme that exchanges free queuosine with guanosine (Yokoyama and Nishimura, 1995).

Some tRNA modifications are general. For example, nearly all tRNAs have the dihydro-uridine (D) in D-loops or ribothymidine in T-loops and most tRNAs have an hypermodified nucleoside such as wyosine next to the anticodon (Bjork et al., 1999). Moreover, the most direct effect of modification is seen in the anticodon. Inosine (originated by deamination of adenine) is often located at the first position of the anticodon, where it is capable of pairing with any one of three

bases: U, C, and A (wobble position). This wobble hypothesis explains much of the degeneracy of the genetic code (Muramatsu et al., 1988).

Those modifications of tRNA stabilize its tertiary structure, increase the specificity of the tRNA for its cognate AARS, increase the surface exposure of the tRNA, increase the interaction with initiation factors and elongation factors and are involved in mRNA decoding in the ribosome (Márquez V. and Nierhaus K.H., 2004).

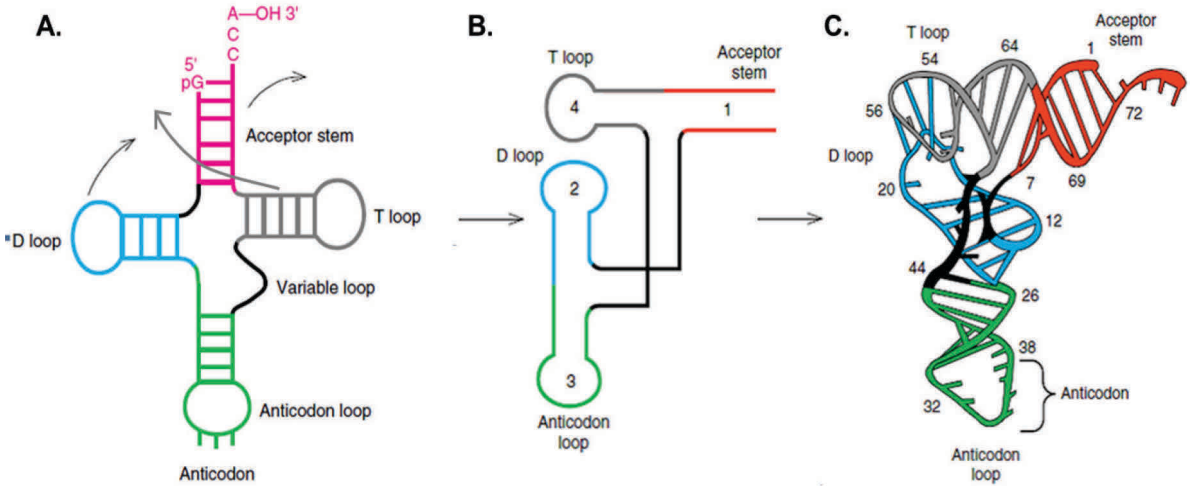


Figure 1-3. Three views of the structure of tRNA. A. General representation for tRNA secondary (cloverleaf) structure. The basic architecture consists of an acceptor stem at the top (magenta), where the amino acid binds to the 3′-terminal adenosine. On the left is the D loop (blue), which contains at least one dihydrouracil base. At the bottom is the anticodon loop (green), containing the anticodon. The T loop (right, gray) contains the invariant sequence TψC. Each loop is defined by a base-paired stem of the same color. **B.** A planar projection of the three-dimensional structure of tRNA. Stackings between the anticodon stem loop and D-stem loop (anticodon arm) and between the acceptor stem and the T-stem loop (acceptor arm) transform the cloverleaf secondary structure of a tRNA into a two domain structure that include an angle of about 90°. **C.** The canonical L-shape tRNAs. This precise architecture allows codon–anticodon interaction at the decoding center. Adapted from Weaver, 2001.

1.1.2.3 Aminoacylation reaction

Each synthetase recognizes a specific amino acid and the corresponding isoacceptor tRNAs, therefore establishing the amino acid–trinucleotide relationship of the code. Aminoacylation of a tRNA proceeds through a two-step mechanism where both steps take place in the active site of the enzyme (Figure 1-4). The first one takes place in the absence of tRNA, with the exceptions of GlnRS, GluRS and ArgRS whose activity is tRNA dependent (Ibba et al., 1997a). At this point, each amino acid fits into an active-site pocket in the AARS (the binding is accomplished via a network of hydrogen bonds, and electrostatic and hydrophobic interactions) and is then activated by attacking a molecule of ATP at the α-phosphate, giving rise to a mixed anhydride intermediate, aminoacyl-adenylate, and inorganic pyrophosphate. Also, ATP is complexed with Mg^{2+} which acts as an electrophilic catalyst by binding to γ- and β-phosphates. An amino acid-activated carboxylic group bound to AMP is an easy target for nucleophilic attack by the ribose 2'- or 3'- OH group in the terminal adenosine of tRNA (Figure 1-4) (Beuning and Musier-Forsyth, 1999).

In the second step, the tRNA first binds the AARS (either as a free enzyme or as an enzyme–aminoacyl-adenylate complex) with recognition depending on sequence-specific protein–RNA interactions. After tRNA binding, the temporary protein–RNA complex catalyzes attachment of the amino acid to the 2'- or 3'- OH terminal adenosine of the tRNA, yielding the specific aminoacyl-tRNA and AMP (Figure 1-4). The mode of the amino acid attachment differs between the two classes of AARS: class I enzymes act on the 2'- OH group of the tRNA, whereas class II enzymes prefer the 3'- OH group. Yet, the aminoacyl group attached to the 2'- OH group of the terminal ribose cannot be used during protein synthesis. It must be later transferred by a transesterification reaction from the 2'- OH to the 3'- OH group of the tRNA terminal adenosine (Márquez V. and Nierhaus K.H., 2004;Antonellis and Green, 2008).

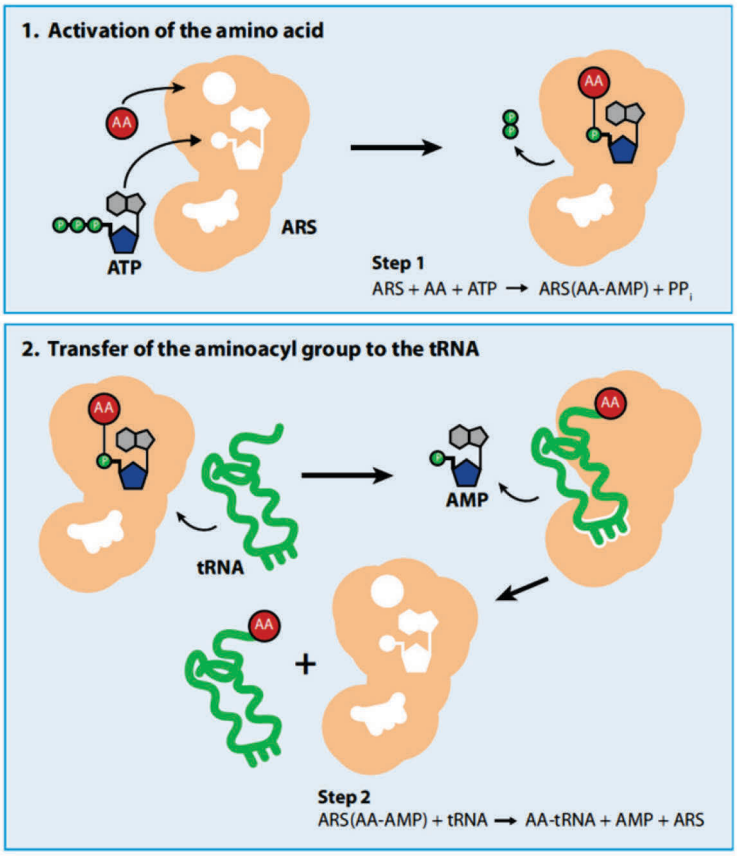


Figure 1-4. The two-step aminoacylation reaction. Each AARS charges a specific tRNA with its cognate amino acid (AA). First, the AARS binds the AA and an ATP molecule to form the aminoacyl-adenylate (AA-AMP), releasing a pyrophosphate molecule (PPi). Secondly, the tRNA binds the AARS and the AA is transferred to the tRNA. An AMP molecule is then released, followed by the charged tRNA. Adapted from Antonellis and Green, 2008.

1.1.2.4 AARS:tRNA interactions

In the translation process, ribosomes do not discriminate among amino acids attached to tRNAs, they only recognize the tRNA part of an aminoacyl-tRNA (Chapeville *et al.*, 1962). Therefore, the responsibility for linking specific amino acids with specific codons (through anticodon-bearing tRNAs) lies with AARS. Since mischarging of tRNAs can lead to the incorporation of the wrong amino acid

into the polypeptide (which can lead to production of inactive or toxic proteins), AARS must be very specific for the tRNAs and amino acids they bring together.

For a particular AARS to recognize the isoaccepting tRNAs of each amino acid, the latter must carry identical signals for this specific recognition. These elements are defined as the “tRNA identity set”, a sum of properties that enable recognition through cognate AARS and prevent binding of non-cognate AARS (Giege, 2006). Figure 1-5 shows a survey of the currently known identity elements described by Giege and colleagues (1998). Most of the class I enzymes (with the exception of LeuRS) require identity elements at the anticodon for proper recognition of the tRNA, while some class II enzymes (SerRS and AlaRS) recognize only features in the acceptor stem (Giege *et al.*, 1998). AARS can employ sophisticated proofreading mechanisms, but since they do not always recognize the anticodon specifically (Ling *et al.*, 2009), as in the cases of Ser, Leu and Ala acceptor tRNAs (Beuning and Musier-Forsyth, 1999), the meaning of a particular codon can be easily changed by mutation of the tRNA. In this context, the universal genetic code is malleable (Schultz and Yarus, 1996).

In these cases, where the anticodon is not important for aminoacylation, Ribas and Schimmel (2001) proposed that the set of nucleotide determinants act as an operational RNA code. These AARS:tRNA interactions relate specific RNA sequences/structures in acceptor stems to specific amino acids, thus, being a sort of “precursor” to the present genetic code. Indeed, the operational RNA code might be a testimony of earlier times in genetic code evolution, when RNA stem-loop structures (precursors of modern tRNAs) were aminoacylated by ribozymes in primitive peptide synthesis mechanisms. During the evolution of the genetic code, these amino acid recognition functions were gradually shifted to codon-anticodon interactions. In fact, the so-called “operational RNA code’’ might have limited the number of variations in tRNA recognition by AARS and these limitations are believed to shape the interpretation of the universal genetic code (De Pouplana and Schimmel, 2004;Shaul *et al.*, 2010).

Besides the anticodon, other significant recognition elements are the discriminatory base at position 73 (tRNAs for chemically similar amino acids have the same nucleoside at position 73), the most distal base pairs of the acceptor

stem and elements at the variable arm of tRNAs. Altogether, 40 positions have been detected as sites for identity signals (Giege, 2006).

Other interesting cases are the so-called “negative determinants”. They do not improve the recognition by cognate AARS but prevent recognition by non-cognate synthetases and thus mischarging (Giege *et al.*, 1998). Such antideterminants can be modified or unmodified nucleotides at any structural domain of the tRNA. Several examples are known, but one is of special interest – the Leu/Ser recognition system, where the A73 protects the $tRNA_{Leu}$ against the SerRS, whereas the G73 protects the $tRNA_{Ser}$ against the LeuRS (Breitschopf *et al.*, 1995;Soma *et al.*, 1996).

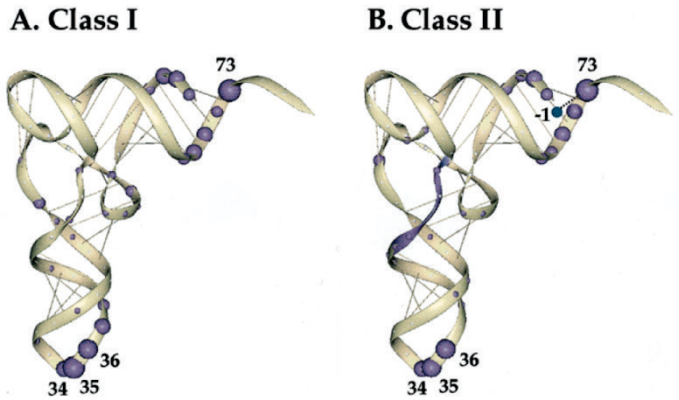


Figure 1-5. **Distribution of identity elements over the tRNA structure.** The tRNA identity elements are distributed as follows: the discriminator base 73, the acceptor stem and the anticodon-loop. The involvement of each feature in tRNA recognition by either class I or class II AARSs is indicated. Apart from these, the variable arm is a key player for Ser identity. Adapted from Giege *et al.*, 1998.

As a final point, studies on the conservation of aminoacylation among different species from all kingdoms revealed contradictory results. Despite the fact that the genetic code is almost universal, there are many examples of species-specific aminoacylation reactions. Cross-aminoacylation experiments described by Soma and Himeno showed that SerRS and LeuRS from *E. coli* are not able to aminoacylate cognate yeast tRNAs. On the other hand, yeast enzymes recognize

E. coli cognate tRNAs (Soma and Himeno, 1998). Additionally, prokaryotic and eukaryotic TyrRS, IleRS and GlyRS are not able to recognize the corresponding tRNA species from other organisms. In several cases, cytoplasmic AARS cannot aminoacylate the corresponding mitochondrial or chloroplastic tRNA *in vitro*, which is in part due to the natural variation in the genetic code of these organelles (De Pouplana and Schimmel, 2004) (see 1.3.2).

1.1.2.5 Quality control in tRNA charging

Aminoacylation is the first step in which errors in translation can occur at a significant rate (the error rate in amino acid and tRNA selection is in the order of 10^{-4}-10^{-5} and 10^{-6}, respectively). Of the 20 canonical AARS, half are unequivocally selective. The other ten (MetRS, IleRS, LeuRS, ValRS, AlaRS, LysRS, GlnRS, ProRS, PheRS, and ThrRS) are not as selective and produce incorrect AA-AMPs (at rates 100–1000 slower than the correct AA-AMP) and even mischarge tRNAs (at slower rates compared with correct AA-tRNA). Most noteworthy examples are the synthesis of Ile-tRNA^{Val} by ValRS, Cys-tRNA^{Pro} by ProRS, and Met-tRNA^{Lys}, Cys-tRNA^{Lys}, Thr-tRNA^{Lys}, Ser-tRNA^{Lys}, Ala-tRNA^{Lys}, Leu-tRNA^{Lys}, Arg-tRNA^{Lys} by LysRS (Jakubowski, 2011).

To make sure that the correct amino acid is chosen, several quality control mechanisms exist. The first one takes place at the active site of the AARS, where discrimination is achieved by preferential binding of a cognate amino acid to the correct active site, and by exclusion of larger and chemically different substrates (Ling *et al.*, 2009). Misactivation occurs because the 20 naturally occurring amino acids have a limited scope of structural variation. In fact, they share important resemblances: some have the same isosteric shape (Val–Thr and Cys–Ser), others the same molecular volume (Ile–Val and Pro–Cys) or very similar chemistry (Ser–Thr). In some of these cases the formation of noncognate adenylates is inevitable and must be corrected through editing mechanisms.

In this second round of quality control, noncognate adenylates are destroyed by two different pathways: pretransfer editing (hydrolysis of the noncognate aminoacyl-adenylates) or posttransfer editing (hydrolysis of the mischarged tRNA) (Figure 1-6) (Schimmel, 2011). This hydrolytic editing step occurs at a second active site within an editing domain that has been identified for the class I IleRS, ValRS, and LeuRS and for the class II AlaRS, ThrRS, PheRS and ProRS (Guo and Schimmel, 2011). The input of a particular editing pathway to global editing varies among AARSs. Although all editing AARSs hydrolyze noncognate AA-AMPs (pretransfer editing), some do not deacylate AA-tRNA (posttransfer editing) as in the cases of MetRS, LysRS and SerRS (Jakubowski, 2011).

Pretransfer editing can be either tRNA dependent or tRNA independent and occurs through a variety of mechanisms that promote hydrolysis of the noncognate aminoacyl adenylate species, leading to the release of the noncognate amino acid, AMP and PPi. The currently accepted pathways include translocation, selective release, and active hydrolysis (Figure 1-6). Pretransfer editing at the active site is used for the clearance of misactivated Ala by ProRS, Ile by LeuRS, Val by IleRS and Ser by ThrRS. Among these, pretransfer editing is predominant in archeal and mitochondrial ProRS, LeuRS and PheRS that naturally lack posttransfer editing domains (Martinis and Boniecki, 2010).

Post-transfer editing of a noncognate aminoacyl-tRNA requires the 3'-CCA of the tRNA, which is linked to the amino acid, to move from the synthetic active site into the editing site, where the RNA–amino acid ester linkage is hydrolysed. This can occur in *cis* (before the mischarged tRNA is released from the synthetase) or in *trans* (by rebinding of the released mischarged tRNA and ensuing clearance of the noncognate amino acid) (Figure 1-6) (Schimmel, 2011).

Moreover, recent studies demonstrated that editing can also be performed by new enzymes with some degree of similarity with AARS. For example, AlaXp (ubiquitous protein homologous to the editing domain of AlaRS) deacylates Ser-$tRNA^{Ala}$ *in vitro* preventing serine toxicity in *E. coli* harboring editing-defective AlaRS (Chong *et al.*, 2008;Guo and Schimmel, 2011).

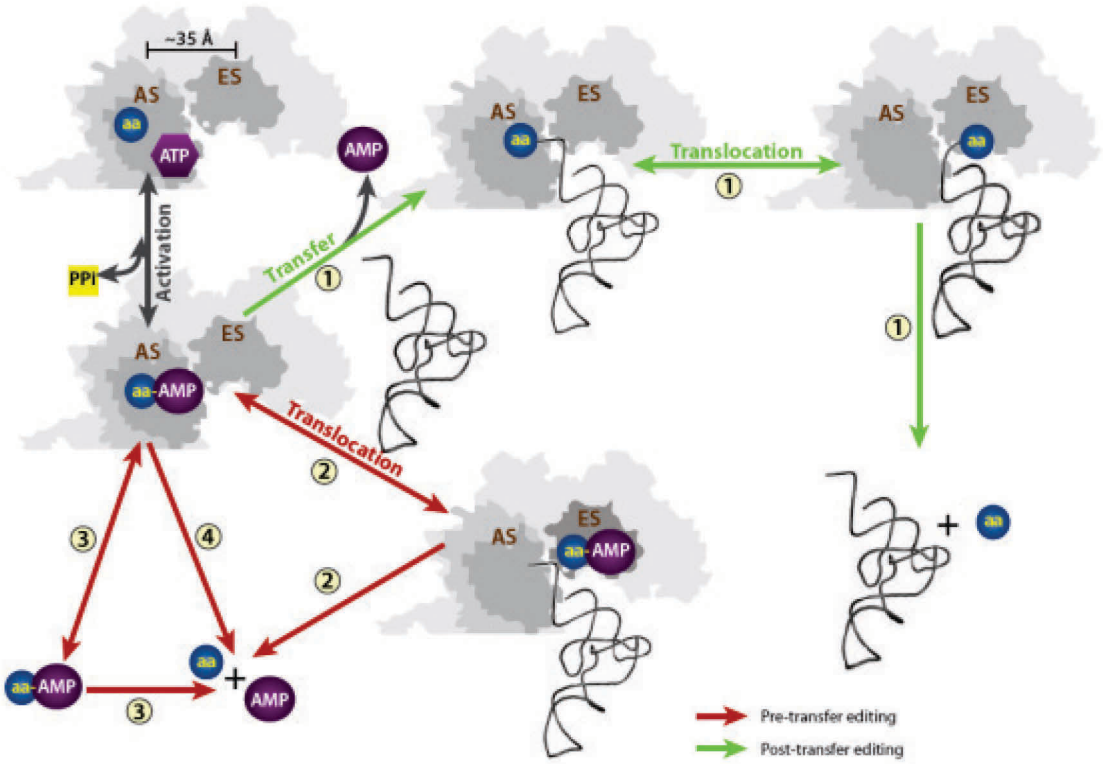


Figure 1-6. Editing pathways against the noncognate amino acid (aa). (Pathway 1) Posttransfer editing. The misactivated aa is first attached to the tRNA and then translocated from the active site (AS) to the distinct editing site (ES) for hydrolysis. (Pathway 2) Translocation of aminoacyl-adenylate (aa-AMP). The misactivated aa-AMP directly translocates to the ES to be hydrolyzed. (Pathway 3) Selective release. After activation, the noncognate but not the cognate aa-AMP is expelled into solution and subjected to spontaneous hydrolysis. (Pathway 4) Active site hydrolysis. The noncognate aa-AMP is hydrolyzed at the AS before release. Pathways 2–4 collectively comprise pretransfer editing. Adapted from Ling *et al.*, 2009.

Since the cellular environment is full of metabolic intermediates and other biogenic amino acids, the extent of discrimination goes further than the 20 canonical amino acids. Evidence for this is homocysteine (metabolic precursor of Met) which can be activated by the MetRS, but is efficiently edited *in vitro* and *in vivo* in a way that the enzyme-bound homocysteinyl adenylate is cyclized to produce homocysteine lactone (Jakubowski, 2011). Other Met analogs (norleucine

and selenomethionine) are recognized in aminoacylation reactions from several organisms, including *E. coli*, and then loaded onto $tRNA^{Met}$. In some of these cases, editing does involve specific features of the amino acid analog, suggesting that this lack of absolute substrate specificity of AARS could be a route for the expansion of the amino acid repertoire *in vivo* (Budisa *et al*., 1999).

1.2 Protein synthesis: the process of mRNA translation

1.2.1 Mechanism of mRNA translation

As mentioned in section 1.1, protein synthesis requires two steps, transcription and translation, functioning with high degree of interdependence, synchronization and efficiency. After mRNA is copied from a DNA strand it is captured by the ribosome where the sequence of codons directs the synthesis of a polypeptide chain. In order to accomplish that, ribosomes work in conjunction with aminoacylated tRNAs and translational factors (Nierhaus K.H. and Wilson D.N., 2004).

The ribosome is a large ribonucleoprotein particle that comprises two subunits in all species. In eukaryotes, ribosomes are in the form of an 80S (svedberg's unit of sedimentation) monosome that comprises a 40S small subunit (that includes the 18S rRNA) and a large 60S subunit (that includes the 28S and 5S rRNAs) (Schmeing and Ramakrishnan, 2009). A common feature between the two subunits is the presence of three binding sites for tRNA: the aminoacyl (A) site accepts the arriving aminoacylated tRNA, the peptidyl (P) site holds the tRNA with the growing peptide chain and the exit (E) site holds the deacylated tRNA before it leaves the ribosome. Both subunits work coordinately to bind the tRNA and the mRNA. The small one binds mRNA and the anticodon loop of the tRNA (monitoring base pairing between codon and anticodon in the decoding process) while the large subunit binds the acceptor arm of the tRNA (catalyzing the peptide bond formation between the incoming amino acid and the nascent peptide chain) (Ramakrishnan, 2002).

Overall, mRNA translation can be divided in initiation, elongation and termination/recycling phases.

1.2.1.1 mRNA translation initiation

Prior to the initiation phase, two important events occur, namely generation of a supply of aminoacyl-tRNAs (reviewed in 1.1.2.3) and the dissociation of ribosomes into their two subunits. This is critical because the assembly of the initiation complex occurs exclusively in the small subunit. The current model for the eukaryotic translation initiation is pictured in Figure 1-7 and involves a series of translation initiation factors.

The first step of the pathway is the assembly of the eIF2•GTP•Met-tRNAi ternary complex (this exchange reaction is facilitated by eIF2B). Then, the complex binds to the 40S ribosomal subunit, forming the 43S pre-initiation complex. This binding is facilitated by eIFs 1, 1A and 3; or eIFs 1, 3 and 5 in yeast (Kapp *et al.*, 2006).

The cap-binding complex (consisting of eIF4E, eIF4G and eIF4A) gathers on the 5'-cap of the mRNA and unwinds secondary structures found in the 5'-untranslated region (by the action of ATP dependent eIF4A and eIF4B factors). Then, eIF4F loads the mRNA onto the 43S complex, in combination with eIF3 and the poly-A binding protein bound to the 3'-poly-A tail, forming the 48S complex. This structure starts scanning the message in the 5'-3' direction, looking for the initiation codon AUG. Once it is found, codon-anticodon base pairing occurs between the AUG codon and the initiator tRNA in the ternary complex. This triggers GTP hydrolysis by eIF2, which is facilitated by eIF5, releasing eIF2•GDP and discharging the Met-tRNAi into the P site of the 40S subunit. It is currently accepted that eIF1, eIF1A, eIF3, and eIF5 also dissociate from the 48S pre-initiation complex at this stage. The circumstance that most factors dissociate from the 40S, together with GTP hydrolysis by eIF5B, facilitates the joining of the 60S subunit to form the 80S monosome. Because the eIF5B•GDP complex has a low

affinity for the ribosome, it dissociates from the complex and this is presumed to be the end of translation initiation (Kapp and Lorsch, 2004;Klann and Dever, 2004).

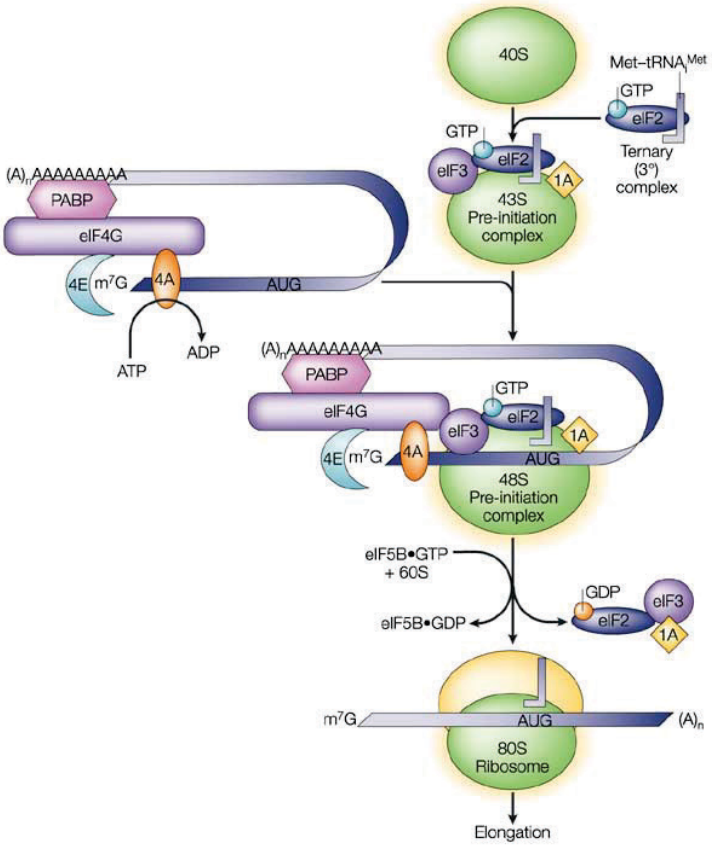


Figure 1-7. Summary of translation initiation in eukaryotes. With the help of eIF2, Met-tRNAi^{Met} binds to the 40S particle, forming the 43S complex. Aided by eIF4, the mRNA binds to the 43S complex, forming the 48S complex. This complex scans in an ATP-dependent reaction in search of a start codon. After pairing of the anticodon of Met-tRNAi with the AUG codon, the eIF5 factor helps the 60S ribosomal particle to bind to the 48S complex, yielding the 80S complex which translates the mRNA. Adapted from Klann and Dever, 2004.

1.2.1.2 mRNA translation elongation

Elongation of a polypeptide chain occurs in three steps that are repeated over and over. Figure 1-8 schematically depicts those steps: occupation of the A-site by the incoming tRNA, peptide-bond formation and translocation (Schneider-Poetsch *et al*., 2010).

Peptide chain elongation begins with a peptidyl tRNA in the ribosomal P-site next to a vacant A-site, so that a second aminoacyl-tRNA can bind to it. Such binding requires a protein elongation factor eEF1A and GTP. Ternary complex eEF1A•GTP•AA-tRNA can be formed with either the cognate or noncognate aminoacyl tRNAs and both can bind to the ribosomal A-site. However, several quality control steps at the decoding and accommodation reactions are used to ensure that only the cognate tRNA is selected for entry into the next stage of elongation (see 1.2.2) (Rodnina and Wintermeyer, 2001a). Correct codon-anticodon base pairing between the mRNA and the tRNA triggers eEF1A GTPase activity. eEF1A•GDP releases the aminoacyl-tRNA into the A-site so that it can continue with peptide bond formation. This occurs at the ribosomal peptidyl transferase center where the nascent polypeptide chain is transferred from the P-site tRNA onto the aminoacyl moiety of the A-site tRNA (Moore and Steitz, 2003). This leaves a deacylated tRNA with its acceptor end in the E-site of the large ribosomal subunit and its anticodon end in the P-site of the small subunit. In turn, the peptidyl-tRNA has its acceptor end in the P-site of the large subunit and its anticodon end in the A-site of the small subunit (Nierhaus K.H. and Wilson D.N., 2004).

Finally, translocation comprises the movement of the mRNA with its peptidyl-tRNA attached (mRNA•tRNA located at the A-site) on the ribosome by one codon length to the left. This is enhanced by eEF2 in combination with GTP hydrolysis. This movement places the deacylated tRNA into the E-site and the peptidyl-tRNA into the P-site, thus relieving the A-site for the next aminoacyl-tRNA. This cycle is repeated until a stop codon is found and the process of termination is initiated (Weaver, 2001;Kapp and Lorsch, 2004).

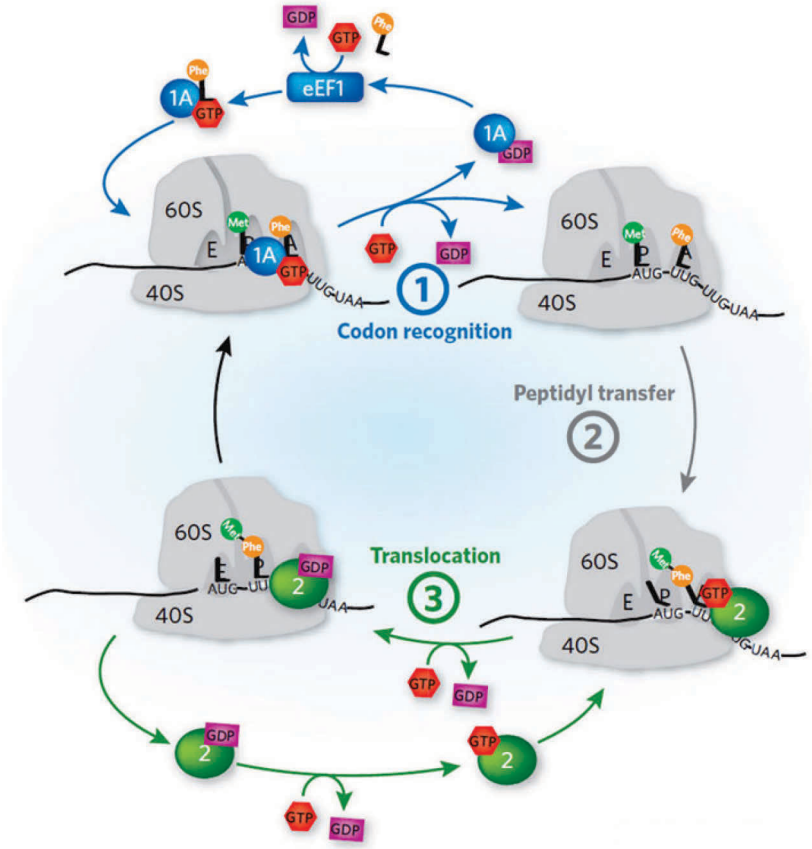


Figure 1-8. Summary of translation elongation in eukaryotes. In the first step, the AA-tRNA forms a ternary complex with eEF1A and GTP, and binds to the A-site of the 80S ribosome. If the tRNA anticodon matches the mRNA, eEF1A hydrolyzes GTP, dissociates from the ribosome and leaves its tRNA behind. In the second step, the ribosome catalyzes peptide bond formation, thus transferring the nascent peptide chain onto the A-site tRNA, and consequently leaving the P-site tRNA deacylated. Finally, eEF2 aids the movement of the A-site tRNA with the peptide chain into the P-site, as well as the movement of the deacylated tRNA into the E-site. Therefore, the A-site becomes available for the next amino-acyl tRNA. Adapted from Schneider-Poetsch *et al*., 2010.

1.2.1.3 mRNA translation termination and recycling

The termination of translation starts as soon as a stop codon on the mRNA encounters the ribosomal A-site (Figure 1-9). In eukaryotes, the termination process is accomplished with the help of the release factors eRF1 and eRF3. The

first binds to the ribosome and recognizes the three stop codons (UAA, UAG and UGA). This triggers the hydrolysis and release of the peptide chain from the tRNA in the P-site. eRF3 is a ribosome-dependent GTPase that helps eRF1 release the finished polypeptide (Kapp and Lorsch, 2004).

After releasing of the polypeptide from the tRNA, the ribosome stays with mRNA and a deacylated tRNA in the P/E hybrid state (the acceptor end is in the E-site of the 60S subunit while the anticodon end is in the P-site of the 40S subunit). Therefore, to prepare for the next round of translation, mRNA is recycled, the deacylated tRNA is released and the 80S ribosome is separated into its two subunits. While in prokaryotes a model for recycling has been proposed (Lancaster *et al.*, 2002), in eukaryotes the exact mechanism of how this happens is not entirely clear.

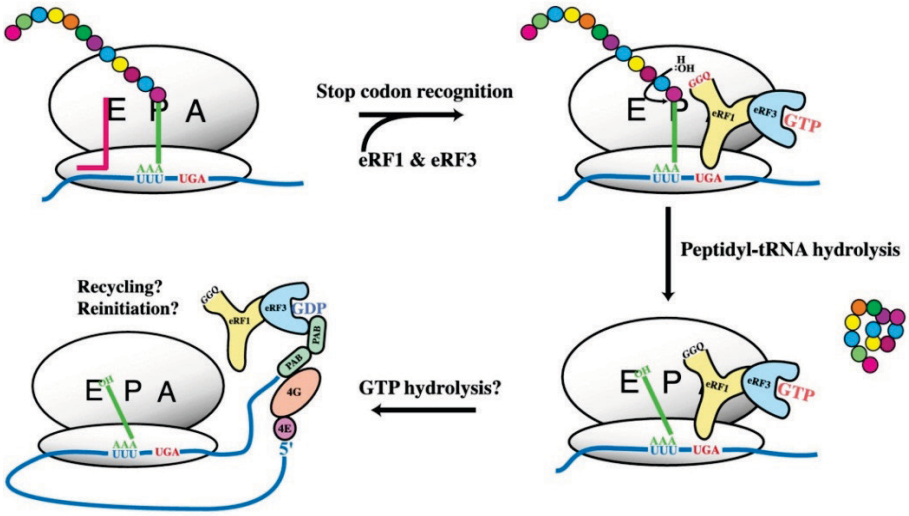


Figure 1-9. Summary of translation termination in eukaryotes. Eukaryotes have two release factors: eRF1, which recognizes the three termination codons and eRF3 which is a ribosome-dependent GTPase that helps eRF1 release the finished polypeptide. Adapted from Kapp and Lorsch, 2004.

1.2.2 Ribosomal proofreading

Global error rates in translation have been estimated to range from 10^{-3} to 10^{-4} *in vivo* in numerous systems, namely *E. coli* and mammalian cells (Kramer and Farabaugh, 2007;Reynolds *et al.*, 2010a). These values are the net accumulation of errors from the sequential steps involved in translation. To minimize the number of errors, two key events are subjected to rigorous quality control or proofreading, namely the synthesis of a correct aminoacyl-tRNA and the stringent selection of the aminoacyl-tRNA by the ribosome.

After an aminoacyl-tRNA is delivered to the ribosomal A-site as part of a ternary complex (eEF1A•GTP•AA-tRNA), the correct codon-anticodon pairing must occur in the decoding center. However, there are a number of competitor tRNAs that complicate the selection process. In bacteria there are 41 tRNAs with different anticodons and in eukaryotes this number is even higher. Among these different tRNAs, approximately 10% are near-cognate, i.e., have an anticodon similar to the cognate tRNA, which poses the question of how the ribosome discriminates on the basis of such a small discrimination area - the anticodon (1kDa) (Wilson and Nierhaus, 2003). To complicate matters further, many tRNAs have post-transcriptional modifications in their anticodons that can expand or restrict *wobble* during codon recognition (Effraim *et al.*, 2009). To overcome these difficulties there are two selection steps in AA-tRNA discrimination, namely a decoding step followed by an accommodation step (Rodnina and Wintermeyer, 2001b). During the first step, the A-site is in a low affinity state, which allows sampling of codon–anticodon interactions based on Watson-Crick base-pairing. Since Watson-Crick base pairs are not significantly stronger than some of the mismatches that would infringe the genetic code (Rodnina, 2012), a second step corrects simple mistakes in the codon-anticodon interaction – the kinetic proofreading mechanism (Figure 1-10).

In this step, the correct geometry of a cognate codon–anticodon helix leads to a series of conformational changes in the ternary complex (eEF1A•GTP•AA-tRNA) that accelerate the rate of GTP hydrolysis by the elongation factor. Conversely, a mismatch in the codon–anticodon pairing does not permit the

conformational changes that are necessary for GTP hydrolysis (Pape *et al.*, 1999;Gromadski and Rodnina, 2004;Ogle and Ramakrishnan, 2005). Once GTP is hydrolysed, the elongation factor releases the AA-tRNA, which can either enter the large subunit A-site or is rejected from the ribosome. If the AA-tRNA bounds to the A-site and forms a peptidyl-tRNA, it is then translocated to the P-site. Here, it is again possible to identify a mismatched codon–anticodon pair since the ribosome also monitors for potential errors in P-site codon–anticodon interactions. Loss of specificity in the A- and P-sites result in an amplification of errors and the premature termination of elongation and accelerated peptide release (Zaher and Green, 2009).

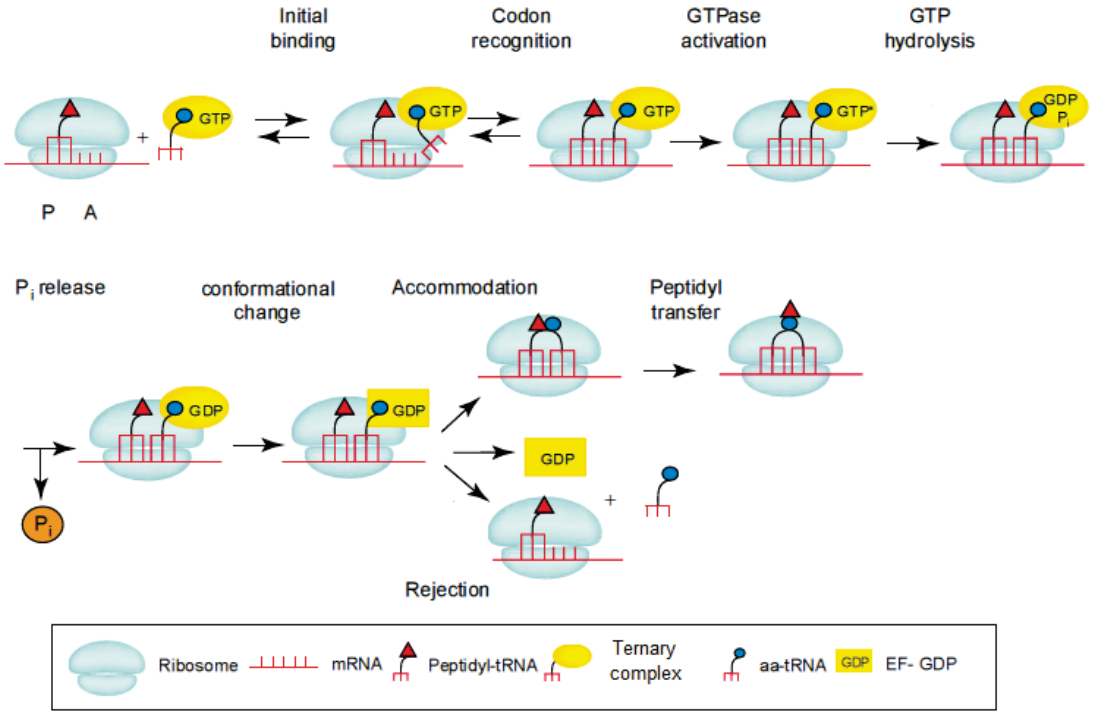


Figure 1-10. Quality control in the ribosome. Kinetic mechanism of AA-tRNA binding to the A site of the ribosome. The binding sites of the ribosome for peptidyl-tRNA and aminoacyl-tRNA are designated P and A, respectively. Elongation factor is depicted differently in the GTP- and GDP-bound conformations, GTP* denotes the GTPase state. Abbreviation: Pi, inorganic phosphate. Adapted from Rodnina and Wintermeyer, 2001b.

In addition to monitoring codon–anticodon interactions in the A- and P-sites, the ribosome also has to maintain the correct mRNA reading to prevent frameshift errors (Reynolds *et al.*, 2010a). Loss of the correct reading frame usually results in out-of-frame truncated proteins. Finally, the ribosome must avoid processivity errors that involve readthrough of nonsense codons and premature termination that leads to production of truncated proteins. Overall, the most common errors during protein translation are frameshifts, premature truncations and misincorporations (Wilson and Nierhaus, 2003).

1.2.3 mRNA mistranslation

The fidelity of protein synthesis is guaranteed by the above mentioned quality control mechanisms. The importance of such mechanisms is highlighted in Figure 1-11, where the potential cellular impact of mistranslation is depicted. Errors in mRNA decoding can impair the fitness of an organism by reducing the amount of functional proteins, increasing the number of toxic proteins, or increasing protein misfolding, which overloads the chaperone machinery and consumes the energy resources of the cell (Nangle *et al.*, 2006). Indeed, some microorganisms possess strict quality control mechanisms to prevent some of these effects. However, proofreading is fallible and errors can occur, leading to the formation of mistranslated proteins.

Studies carried out in the Schimmel's laboratory link mistranslation (originated from defective editing by a tRNA synthetase) with diseases arising from misfolded protein aggregates and autoimmunity (Nangle *et al.*, 2006;Schimmel, 2008a;Schimmel, 2008b). Investigations with a mammalian cell culture system with editing-defective ValRS showed that this enzyme misactivates threonine and catalyzes production of Thr-tRNA^{Val} which is used in protein synthesis. The erroneous decoding decreases cell viability and induces an apoptotic response (Nangle *et al.*, 2006). Also, a single mutation in the editing domain of the mouse AlaRS (Ala734Glu) induces mischarging of alanine tRNAs with serine. The

presence of Ser-tRNA^{Ala} in the cell promotes global misincorporation of serine at alanine codons causing protein misfolding, aggregation and ubiquitination. In turn, this activates autophagy and the unfolded protein response, triggers rapid loss of Purkinje cells and mouse premature death (Lee *et al.*, 2006). The MetRS is one of the most peculiar cases since it mischarges methionine tRNAs with homocysteine in a wide variety of organisms. Misincorporation of homocysteine at methionine codons has been described for bacteria, yeast and mammalian cells. In humans, it causes proteome N-homocysteinylation in vascular endothelial cells (HUVEC) and increases the risk of vascular disease (Jakubowski, 2008;Glowacki *et al.*, 2010).

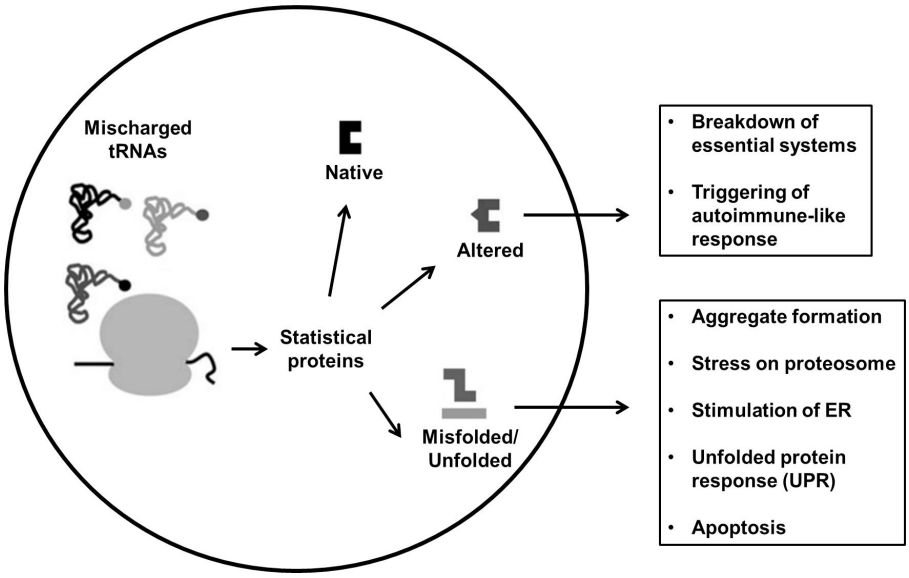


Figure 1-11. Cellular impact of global mistranslation. Mistranslation resulting from the release of mischarged tRNAs in an eukaryotic cell can alter a series of essential systems. For example, when the production of unfolded proteins exceeds the capacity of the cell to remove them, toxic aggregates can form. Adapted from Nangle *et al.*, 2006.

Despite the unequivocal importance of quality control mechanisms, there are cases where a certain level of errors during translation can be tolerated or even generate advantageous phenotypes. Therefore, quality control mechanisms are sometimes dispensable for viability (Santos *et al.*, 1999;Pezo *et al.*, 2004;Bacher *et al.*, 2007;Ruan *et al.*, 2008;Meyerovich *et al.*, 2010;Reynolds *et al.*, 2010b). For example, *Mycoplasma spp.* have error-prone LeuRS, PheRS and ThrRS with mutations and deletions in their editing domains (Li *et al.*, 2011). The

mischarging of tRNAs by these AARSs increases mistranslation and leads to the production of statistical proteins with Phe/Tyr, Leu/Met and Leu/Val substitutions. The high frequency of mistranslation may provide a mechanism for the remarkable phenotypic plasticity of *Mycoplasma* that in turn allow it to escape host defences (Li *et al.*, 2011). More clear examples of tolerance to mistranslation include *E. coli* mutants that mistranslate Asp-tRNA^{Asn}, Glu-tRNA^{Gln} or Cys-tRNA^{Pro} more than 10% of the time and do not exhibit a growth defect (Ruan *et al.*, 2008); *Candida albicans* can tolerate up to 28% misincorporation of leucine at CUG positions and uses such ambiguity to generate extensive phenotypic diversity (Gomes *et al.*, 2007;Miranda *et al.*, 2007); and *Acinetobacter baylyi* mutants with an editing defective IleRS that produces Val-tRNA^{Ile} display selective advantages in media with limiting isoleucine and excess of valine (Bacher *et al.*, 2007). Another interesting case that highlights the positive aspects of mistranslation concerns the mammalian MetRS. Under environmental stress, this enzyme misacylates noncognate tRNAs leading to Met incorporation at various non-cognate sites. Met is a scavenger of reactive oxygen species (ROS) and consequently provides a response mechanism against stress (Netzer *et al.*, 2009).

1.3 Development of the genetic code

1.3.1 Origin and early evolution of the genetic code

Unravelling the origin of the genetic code and translation machinery is a difficult problem because the events shaping the genetic code took place long time ago. Despite this fact, three main theories are conceivable based on the available data, namely the Stereochemical Theory (Yarus *et al.*, 2005), the Adaptive Theory (Woese, 1965) and the Coevolution Theory (Wong, 1975).

The modern version of the stereochemical theory posits that the primordial code functioned through interactions between amino acids and cognate triplets that resided within amino acid binding RNA molecules (Yarus *et al.*, 2005). This

precipitated the emergence of the primordial code that subsequently evolved into the modern code in which the specificity is maintained by much more precise and elaborate, indirect mechanisms involving tRNAs and aminoacyl-tRNA synthetases.

The adaptive theory of the code evolution postulates that the structure of the genetic code was shaped under selective forces that made the code maximally robust, that is, minimize the effect of errors on the structure and function of the synthesized proteins (Woese, 1965). The fact that related codons code for similar amino acids and the observation that mistranslation occurs more frequently in the first and third positions of codons, whereas it is the second position that correlates best with amino acid properties were interpreted as evidence in support of the adaptive theory (Woese, 1965). More recently, it has been shown that the nonrandom assignments of amino acids in the standard code can be almost completely explained by incremental code evolution by codon capture or ambiguity reduction processes (Stoltzfus and Yampolsky, 2007).

The coevolution theory hypothesizes that the structure of the standard code reflects the pathways of amino acid biosynthesis (Wong, 1975). According to this scenario, prebiotic synthesis could not produce 20 modern amino acids, so a subset of the amino acids had to be produced through biosynthetic pathways before they could be coopted into the genetic code and translation, and hence coevolution of the code and amino acid metabolism (Wong and Bronskill, 1979). The Coevolution Theory suggests the existence of three critical moments (steps) in the development of the genetic code. An initial step (Phase-1) characterized by the incorporation of the prebiotic amino acids Gly, Ala, Ser, Asp, Glu, Val, Leu, Ile, Pro and Thr. An intermediate step (Phase-2) involving the incorporation of 7 additional amino acids derived from the prebiotic ones through biosynthetic means, namely Phe, Tyr, Arg, His, Trp, Lys and Met. And, a final step (Phase-3) where the five amino acids whose synthesis is tRNA-dependent or is mediated through non-canonical biosynthetic pathways, namely Asn, Gln, Cys, selenocysteine (Sec) and pyrrolysine (Pyl), were incorporated into the genetic code (Wong, 1975;Koonin and Novozhilov, 2009).

None of the three major theories on the origin of the genetic code is indisputably supported by the currently available data. Therefore, it is conceivable that each of these theories captures some aspects of the code's origin.

1.3.2 Natural variations in the universal assignment of codons

The genetic code was thought to be immutable and universal until the late 70's when sequencing of genes from the human mitochondrial genomes revealed deviations in codon assignments when compared with their nuclear counterparts (Barrell *et al.*, 1979). Since then, over 20 alternative codes have been reported (Figure 1-12) (Knight *et al.*, 2001;Santos *et al.*, 2004) and most of them are known from mitochondria that have small genomes (less than a 100 protein-coding genes) (Sengupta *et al.*, 2007). Beyond mitochondrial genomes, code variants are very rare. Despite this, many cases are now known where a codon has been reassigned from one amino acid to another, from a stop to an amino acid, or from an amino acid to a stop.

Naturally occurring variants of the canonical genetic code in mitochondria are shown in Figure 1-12. For example, the standard UGA stop codon codes for tryptophan in the mitochondria of most organisms (Yokobori *et al.*, 2001). The arginine AGA and AGG codons have been reassigned to serine in bilateria, to glycine in urochordates and became stop codons in vertebrates. The isoleucine AUA codon represents a peculiar change since it codes for methionine in most metazoan mitochondria, but in platyhelminths, echinoderms and hemichordates it changed again to the standard isoleucine (Castresana *et al.*, 1998). The leucine CUN codons are decoded as threonine in yeast (Pape *et al.*, 1985) and lysine AAA codons are decoded as asparagine in platyhelminths and echinoderms (Castresana *et al.*, 1998). Finally, the arginine CGN codon family is unassigned in yeast mitochondria (Pape *et al.*, 1985).

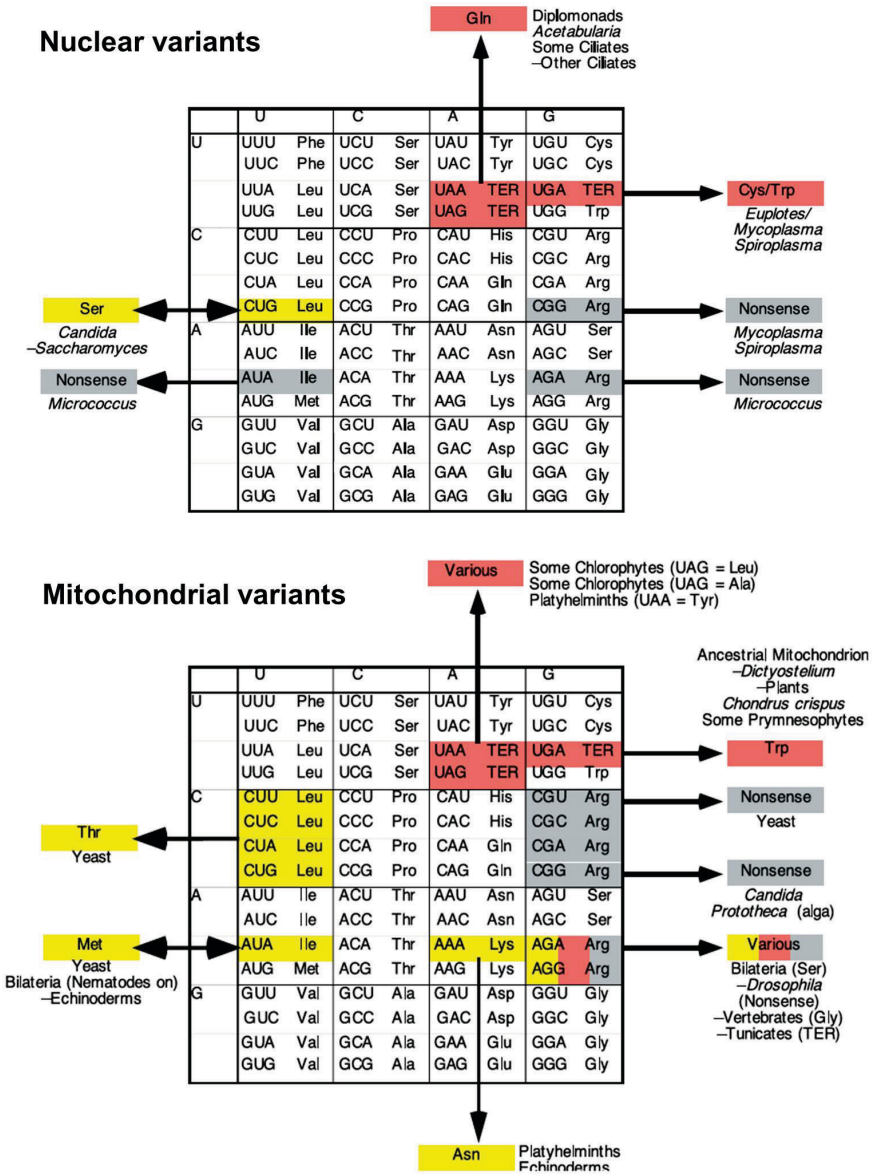


Figure 1-12. Genetic code alterations to the canonical genetic code (nuclear and mitochondrial variants). Yellow concerns missense changes, grey indicates nonsense changes and red indicates changes in termination codons. Adapted from Knight *et al.*, 1999.

All nuclear variants (Figure 1-12) involve codons that have been reassigned in mitochondrial genomes. The genera *Oxytricha*, *Paramecium* and *Tetrahymena* decode UAA and UAG codons as glutamine; three peritrich species also translate the UAA stop as glutamate and the UGA stop has been reassigned to cysteine in *Euplotes* (Lozupone *et al.*, 2001;Sanchez-Silva *et al.*, 2003). Remarkably, in

Bacillus subtilis the UGA stop codon is decoded as tryptophan, but also retained its ability to be used for translation termination (Lovett *et al.*, 1991). Some bacteria of the *Mycoplasma* and *Spiroplasma* genera changed the identity of the UGA stop codon to tryptophan (Oba *et al.*, 1991b), and the arginine CGG codon is unassigned (Oba *et al.*, 1991a). In *Micrococcus*, the arginine AGA and the isoleucine AUA codons are also unassigned (Ohama *et al.*, 1990). Finally, the only documented case of sense codon reassignment in nuclear genes is the decoding of the standard leucine CUG codon as serine in several fungal species of the CTG clade (see 1.4.3) (Tuite and Santos, 1996;Butler *et al.*, 2009).

1.3.3 Natural expansion of the genetic code to 22 amino acids

A wide range of pathways are available for molecular variations between a gene and its corresponding protein (suppression, splicing, cotranslational and posttranslational modifications). Among those pathways, recoding events represent an intriguing way of circumvent the standard translation process in order to produce different protein products (Gesteland and Atkins, 1996). Recoding can be facilitated by the presence of stem-loops, pseudoknots or specific 5' or 3' sequence elements in the mRNA structure that permit decoding phenomena like bypass (Herr *et al.*, 2000), frameshift (Farabaugh, 1996), readthrough (Baranov *et al.*, 2002), reassignment or suppression (Poole *et al.*, 2003). For example, the termination UGA and UAG codons can be used to incorporate noncanonical amino acids like selenocysteine (Sec) (Bock *et al.*, 1991;Allmang and Krol, 2006;Allmang *et al.*, 2009) and pyrrolysine (Pyl) (Srinivasan *et al.*, 2002;Krzycki, 2005;Gaston *et al.*, 2011b), respectively. These alterations are usually considered by current literature as natural expansions of the genetic code.

Incorporation of Sec in response to an in-frame UGA codon is achieved by a complex recoding machinery to inform the ribosome not to stop at this position on the mRNA. The mechanism is distinct in prokaryotic and eukaryotic organisms, but there are some similarities. For example, both have a special tRNA for Sec,

which is a minor isoacceptor of a serine tRNA. The other key players are SelB and SECIS (selenocysteine insertion sequence). Since Sec has its own $tRNA^{Sec}$, biosynthesis begins, precisely, with SerRS acylating $tRNA^{Sec}$ with serine, producing $Ser\text{-}tRNA^{Sec}$. Then different enzymes convert $Ser\text{-}tRNA^{Sec}$ into $Sec\text{-}tRNA^{Sec}$: selenocysteine synthase (SelA) and selenophosphate synthetase (SelD) in bacteria, and O-phosphoseryl-tRNA kinase (PSTK) and Sep-tRNA:Sec-tRNA synthase (SepSecS) in archaea and eukarya (Figure 1-13) (Ambrogelly *et al.*, 2007;Yuan *et al.*, 2010). Once the $Sec\text{-}tRNA^{Sec}$ is available, recoding of UGA as Sec requires the presence of the translation elongation factor SelB. This factor binds to $Sec\text{-}tRNA^{Sec}$ and forms the SelB.GTP.$Sec\text{-}tRNA^{Sec}$ complex that is delivered to the ribosome. Studies performed by Bock and co-workers revealed that SelB must be complexed with the SECIS element for the correct interaction with the ribosome to occur (Bock *et al.*, 1991). Binding of the ternary complex to the SECIS structure induces a conformational change in SelB that enables the codon-anticodon interaction between the $Sec\text{-}tRNA^{Sec}$ and the UGA codon at the ribosomal A-site to occur. Therefore, the SECIS element has a critical double function. It converts SelB into a "competent state" that gives SelB a strong competitive advantage relative to the release factor for decoding of UGA. Simultaneously, it prevents normal UGA termination codons from being decoded as Sec by the SELB.GTP.$Sec\text{-}tRNA^{Sec}$ ternary complex. Thus, the dual properties of SelB and SECIS ensure that only UGA codons in selenoprotein mRNAs are recoded (Allmang and Krol, 2006).

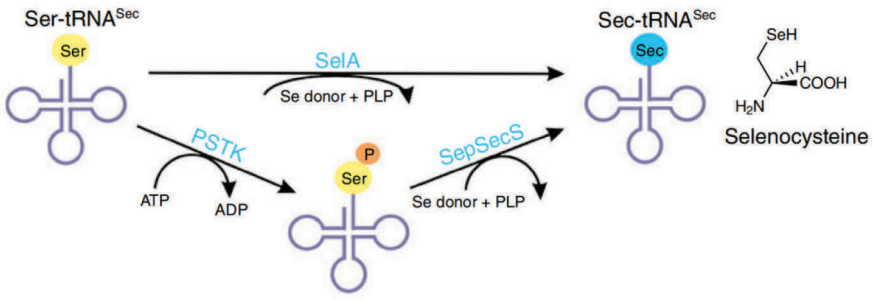


Figure 1-13. tRNA-dependent amino acid transformations leading to selenocysteine. The SelA route is the bacterial pathway; the PSTK/SepSecS pathway is found in archaea and eukaryotes. Adapted from Ambrogelly *et al.*, 2007.

While Sec is generated by a pretranslational modification of Ser-tRNA^{Sec}, pyrrolysine (Pyl) is directly attached to tRNA$^{Pyl}_{CUA}$ by PylRS in response to an in-frame UAG codon in the *Methanosarcina barkeri* monomethylamine methyltransferase gene (Krzycki, 2005). These are methane-producing organisms and Pyl is necessary for methane biosyntesis from methylamines. Indeed, the three different methyltransferases that initiate methanogenesis from different methylamines have genes with an in-frame UAG codon which is translated as the noncanonical pyrrolysine (Srinivasan *et al.*, 2002;Gaston *et al.*, 2011a). The mechanism for Pyl insertion requires a tRNA^{Pyl} (tRNA$^{Pyl}_{CUA}$) and a pyrrolysyl-tRNA synthetase (PylRS) (Figure 1-14). The PylRS is considered the 21st AARS, since it charges specifically Pyl to tRNA$^{Pyl}_{CUA}$ (lysine itself and its cognate tRNA^{Lys} are not substrates of this enzyme) (Blight *et al.*, 2004). Therefore, PylRS is the first example of a synthetase that is specific for a modified amino acid, and PylRS and tRNA^{Pyl} are a naturally occurring AARS-tRNA pair that is effectively orthogonal to the canonical genetic code (Gaston *et al.*, 2011a).

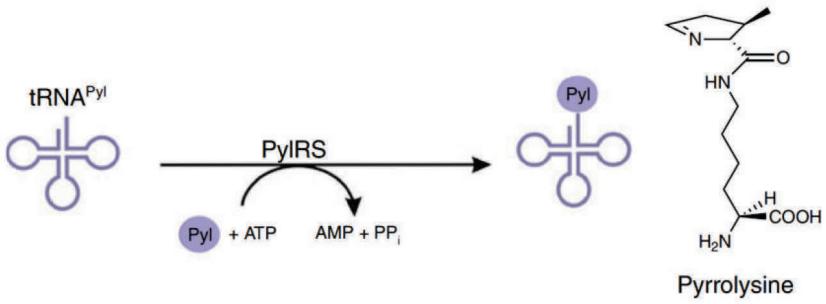


Figure 1-14. Schematic representation of Pyl-tRNAPyl formation. Chemical structures of lysine and pyrrolysine are shown on the side. Adapted from Ambrogelly *et al.*, 2007.

Several mechanisms for Sec and Pyl insertion in protein sequences are present in different organisms, but context dependency is the universal feature of these occurrences and they can be regarded as preprogrammed modifications of canonical decoding rules.

1.3.4 Theories for the evolution of the genetic code

In the light of the variations described above, numerous theories have been formulated to explain the evolution of genetic code alterations, but two main hypothesis are commonly accepted (Figure 1-15) (Knight *et al.*, 2001;Santos *et al.*, 2004;Bacher *et al.*, 2004).

The codon capture theory was proposed by Osawa and Jukes (1989) and its main assumption is that a particular codon cannot have two meanings simultaneously, since ambiguity is detrimental for cell viability (Osawa and Jukes, 1989). Consequently, under biased genome AT or GC pressure, certain codons vanish from the polypeptide coding sequences (ORFeome). These unassigned codons lead to loss of functionality of the corresponding tRNAs, which can be eliminated by natural selection. These erased codons may be reintroduced by genetic drift. When the GC content of the genome changes again, the erased codons re-emerge but they lack cognate tRNAs. Cells that are able to capture these codons and convert them into sense codons have a growth advantage and the codon reassignment can be achieved. The codon capture theory is supported by the disappearance of the CGG codon in *Mycoplasma capricolum* (25% genome G + C) and the AGA and AUA codons in *Micrococcus luteus* (75% genome G+C) (Osawa *et al.*, 1992). On the other hand, there are several other examples of codon reassignments in organisms where GC bias is not visible, and even cases of codon reassignments that appear against such bias. For example, reassignment of the leucine CUU and CUA codons to threonine in the AT rich genome of yeast mitochondria (Santos *et al.*, 2004). In these cases, the codon capture theory cannot explain codon reassignment.

Conversely, the ambiguous intermediate theory proposed by Schultz and Yarus (1996) does not involve the complete disappearance of a codon before its reassignment (Schultz and Yarus, 1996). This theory postulates that ambiguous codon decoding provides an initial step for gradual codon identity change, and wild-type or mutant misreading tRNAs are the critical elements of codon

reassignment. The basic idea is that the appearance of mutant tRNAs with altered/expanded decoding properties are able to recognise and translate non-cognate codons. Consequently, noncognate amino acids are incorporated into proteins in competition with cognate ones and statistical proteins are produced. If this ambiguous codon translation is advantageous for the organism, the alternative codon interpretation could be selected by natural selection, leading to a new arrangement of the code (Miranda *et al.*, 2006). Numerous experimental evidences obtained from engineered codon ambiguity in *E. coli* and yeast support this theory. *E. coli* tolerates up to 10% of ambiguity (Ruan *et al.*, 2008). Under specific conditions, increased levels of ambiguity can even be advantageous as in the case of *Acinetobacter baylyi* (Bacher *et al.*, 2007). Also, in certain *Candida* species the CUG codon is decoded as both serine and leucine, due to charging of a $tRNA^{Ser}_{CAG}$ by both SerRS and LeuRS (Suzuki *et al.*, 1997). Remarkably, cells are highly tolerant to codon ambiguity and trigger a unique stress response that exponentially increases pre-adaptation potential (Santos *et al.*, 1997;Santos *et al.*, 1999;Silva *et al.*, 2007;Moura *et al.*, 2009).

An universal mechanism for the evolution of the genetic code has not yet been formed, however an unifying model proposed by Sengupta and Higgs (2005) postulates that genetic code alterations evolve through a process of gain and loss (Sengupta and Higgs, 2005). The loss can be the deletion or loss of function of a tRNA or release factor. The gain can be the gain of a new type of tRNA or the gain of function of an existing tRNA due to mutation or base modification. According to this hypothesis, mechanisms are favored depending on whether the codon is erased from the genome during reassignment and by the order of the gain and loss events.

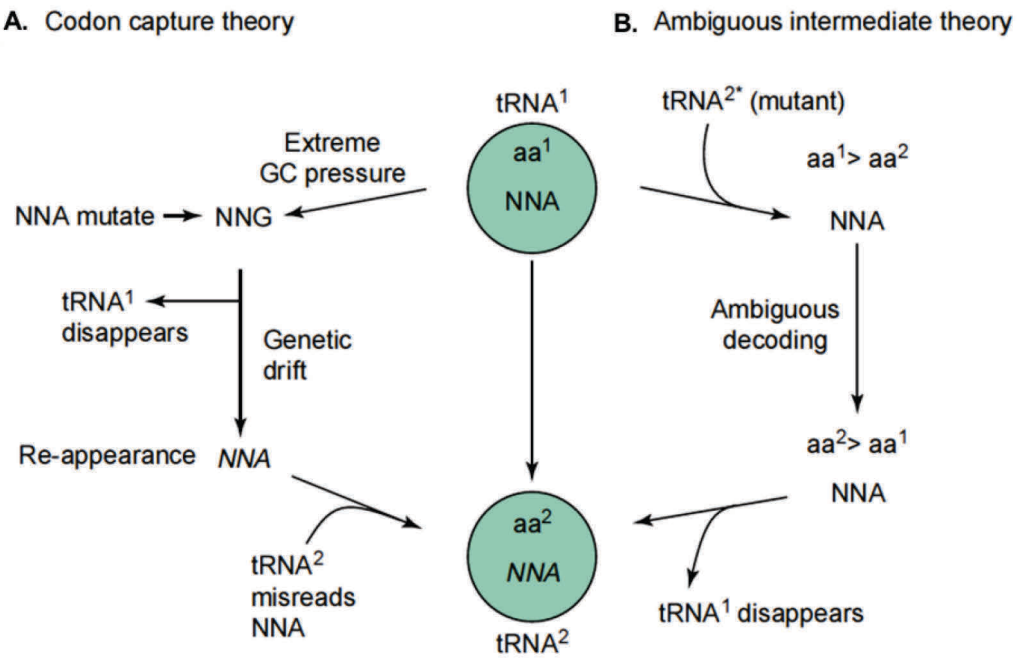


Figure 1-15. Molecular mechanisms of evolution of genetic code alterations. A. The codon capture theory. Evolutionary changes in the base composition (AT-GC) of the entire genome would result in an alteration of codon usage. In extreme cases of bias, a particular codon (NNA) might become unassigned along with the corresponding tRNA (tRNA1). If the unassigned codon reappears by genetic drift, it can be decoded through misreading by a noncognate tRNA (tRNA2), thus changing the identity of the NNA codon to aa2. **B.** The ambiguous intermediate theory. In this case, a mutant tRNA from a non-cognate codon family (tRNA2*) competes with the cognate tRNA (tRNA1) for codon NNX, leading to translational ambiguity of the NNA codon. Eventually, the gene encoding tRNA1 disappears from the genome, while tRNA2 is maintained (exclusively decoding the old NNA codon as aa2). This codon decoding with new meaning is tolerated by the proteome. The mutant tRNA captures new NNA codons that arise by genetic drift, increasing usage of the new NNA codon over time. This is a dynamic process driven by selection. Adapted from Santos *et al.*, 2004.

1.3.5 Variation in codon assignment through codon identity engineering

Despite the variations to the standard code that have been discovered in numerous organisms, the engineering of codon reassignments has been extremely difficult to achieve. Many experimental attempts to expand the code have been reported (Wang and Schultz, 2001;Wang *et al.*, 2006a;Liu and Schultz, 2010;Hoesl and Budisa, 2012), in particular to incorporate chemically synthesized non-canonical amino acids (ncAAs) into recombinant proteins (Figure 1-16).

For isostructural ncAAs (recognized by the endogenous host cell machinery), replacement of canonical amino acids (cAAs) has been carried out using a supplementation-based incorporation method (SPI). This approach uses auxotrophic strains for one of the common 20 cAAs to replace this amino acid with a ncAA. The method exploits the natural tolerance of the endogenous host AARSs to the ncAAs, which allows the concurrent exchange of many residues in a target protein by sense-codon reassignment (Link *et al.*, 2003). Although the overall replacement of a canonical amino acid by a ncAA cannot tolerate exponential growth, nondividing cells are still viable and are able to overexpress proteins that contain the ncAA. The diversity of amino acid analogs that can be incorporated using this approach has been increased by AARS overexpression, active-site engineering and editing domain mutations. Numerous examples of applications of this technique are available. For example, methionine has been replaced with selenomethionine to introduce a heavy atom into proteins for crystallographic phasing experiments and in other cases methionine or phenylalanine have been replaced by alkyne-containing ncAA analogs to track newly synthesized proteins. The down side of this method is the lack of site-specificity (Liu and Schultz, 2010).

As for orthogonal ncAAs (that do not participate in conventional translation), they have been added by site-specific incorporation in response to stop or quadruplet codons (stop codon suppression, SCS) using orthogonal aminoacyl-tRNA synthetase:tRNA pairs (Wang and Schultz, 2001).

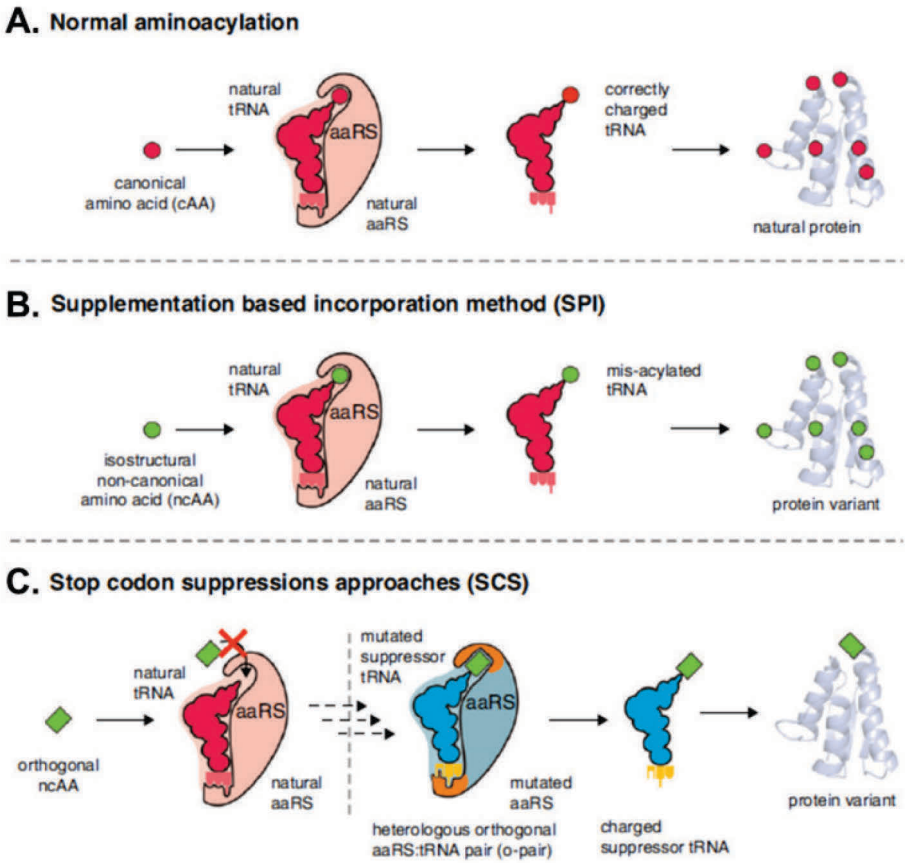


Figure 1-16. Different approaches for the incorporation of unnatural amino acids into proteins *in vivo*. A. In the natural scenario, tRNA aminoacylation is catalyzed by the corresponding aminoacyl-tRNA synthetase (aaRS) responsible for charging the tRNA with the cognate amino acids. **B.** The supplementation based incorporation method (SPI) exploits the natural substrate tolerance of the endogenous host aaRSs by using auxotrophic host strains. The substrates for this procedure are noncanonical amino acids (ncAAs) which are isostructural to their canonical counterparts. **C.** Stop codon suppression methodologies (SCS) are site-specific and make use of a heterologous orthogonal aaRS:tRNA pair to incorporate an orthogonal amino acid in response to a stop or quadruplet codon. Adapted from Hoesl and Budisa, 2012.

Orthogonal tRNAs and AARSs are constructed by obeying to a series of conditions that contribute to the lack of cross-reactivity between the pair and the endogenous host synthetases, amino acids and tRNAs. Firstly, the tRNA cannot be recognized by the endogenous AARSs of the host, but must function efficiently in translation. Another crucial requirement for the tRNA is that it must deliver the

ncAA in response to a unique codon that does not encode any of the 20 cAA (for example, a stop codon). Secondly, the orthogonal AARS must aminoacylate only the orthogonal tRNA and none of the endogenous tRNAs. This synthetase must also aminoacylate the tRNA with only the desired unnatural amino acid and no endogenous amino acid. Similarly, the ncAA cannot be a substrate for the endogenous synthetases. Finally, the ncAA must be efficiently transported into the cytoplasm when added to the growth medium, or biosynthesized by the host (Wang *et al.*, 2006a). A number of heterologous AARS/tRNA pairs have been developed to expand the genetic code of *E. coli*, yeast and mammalian cells. For example, the *E. coli* GluRS/human initiator tRNA, the *E. coli* TyrRS/*E. coli* $tRNA^{Tyr}$, the *E. coli* LeuRS/*E. coli* $tRNA^{Leu}$, and the *M. mazei* PylRS/*M. mazei* $tRNA^{Pyl}$ pairs are orthogonal in *S. cerevisiae* (Liu and Schultz, 2010), demonstating the potential of this methodology for synthetic biology, in particular to solve industrially relevant bio-production problems.

1.4 Evolution of the genetic code in yeast

Numerous variations to the standard code have been found in yeast. The majority of these alterations strongly support the ambiguous intermediate theory (Schultz and Yarus, 1996), in particular because they involve structural alterations of the translational machinery and are driven by selection. Among the alterations discovered so far are the reassignment of the mitochondrial UGA-stop codon to tryptophan in most yeasts, the four leucine CUN (N = any nucleotide) codons that are decoded as threonine in the mitochondria of *Saccharomyces cerevisiae* and *Candida glabrata*, and the reassignment of the leucine CUG codon to serine in the nuclear genes of CTG clade species (Miranda *et al.*, 2006;Butler *et al.*, 2009).

1.4.1 Reassignment of the CUN codon family in yeast mitochondria

Structural alterations in a single mitochondrial tRNA are associated to the change of identity of leucine CUN codons to threonine in *S. cerevisiae* mitochondria. This change results from the loss of the leucine tRNA ($tRNA^{Leu}_{UAG}$) that would translate CUN codons and from the appearance of an atypical threonine tRNA ($tRNA^{Thr}_{UAG}$) (Macino *et al.*, 1979;Li and Tzagoloff, 1979). This mutant $tRNA^{Thr}_{UAG}$ derived from a $tRNA^{His}$ is able to decode the four leucine CUN codons. Its structure is abnormal because it has 6 base pairs in the anticodon stem instead of the canonical 5, and 8 nucleotides in the anticodon loop instead of the canonical 7 (Su *et al.*, 2011). In addition to the atypical $tRNA^{Thr}_{UAG}$, yeast mitochondria also express a standard $tRNA^{Thr}_{UGU}$ that reads the ACN threonine codon box. This $tRNA^{Thr}_{UGU}$ has a typical anticodon stem (5 base pairs) and anticodon loop (7 base pairs) which is in marked contrast with the atypical $tRNA^{Thr}_{UAG}$ (Osawa *et al.*, 1990;Su *et al.*, 2011). Additionally, this alteration to the standard code required a novel synthetase (ThrRS) to aminoacylate the unusual $tRNA^{Thr}_{UAG}$, which confirms that CUN reassignment occurred through structural changes of both the $tRNA^{Thr}$ and its ThrRS (Pape *et al.*, 1985).

1.4.2 Reassignment of the UGA stop codon in yeast mitochondria

Alteration of the translational machinery also mediates the reassignment of the UGA stop codon to tryptophan in yeast mitochondria (Hatfield and Diamond, 1993). In this case, the release factor (eRF1) lost its ability to recognize the UGA codon through changes in its structure (Inagaki and Doolittle, 2001). This alteration permitted the capture of the UGA stop by a near-cognate tryptophan tRNA. However, efficient decoding of UGA as tryptophan required specific mutations in the anticodon of the primordial $tRNA^{Trp}_{CCA}$, namely a C to U mutation in the first position of the anticodon to produce the 5'-UCA-3' anticodon. In this case, the U at the first position of the anticodon is modified to prevent decoding of the cysteine UGC and UGU codons as tryptophan (Yokobori *et al.*, 2001;Miranda *et al.*, 2006).

1.4.3 Reassignment of the CUG codon in the CTG clade

The reassignment of the leucine CUG codon to serine, in the cytoplasm of numerous *Candida* species, *Pichia stipitis*, *Debaryomyces hansenii* and *Lodderomyces elongisporus* (Sugita and Nakase, 1999), is also mediated through structural alteration of the translation machinery (Santos *et al.*, 1996;Perreau *et al.*, 1999). These species form the so-callled CTG clade (Figure 1-17) which is characterized by the only nuclear genetic code change that involves a sense to sense reassignment (Butler *et al.*, 2009). Remarkably, *C. zeylanoides*, *C. dubliniensis*, *C. tropicalis*, *C. guilliermondii*, *C. albicans* and many others translate CUG codons ambiguously (as both serine and leucine), while *C. cylindracea* decodes CUG codons as serine only (Santos and Tuite, 1995;Tuite and Santos, 1996;Suzuki *et al.*, 1997;Gomes *et al.*, 2007). These differences in the reassignment status between organisms of the same clade are related to specificities of the translation machinery.

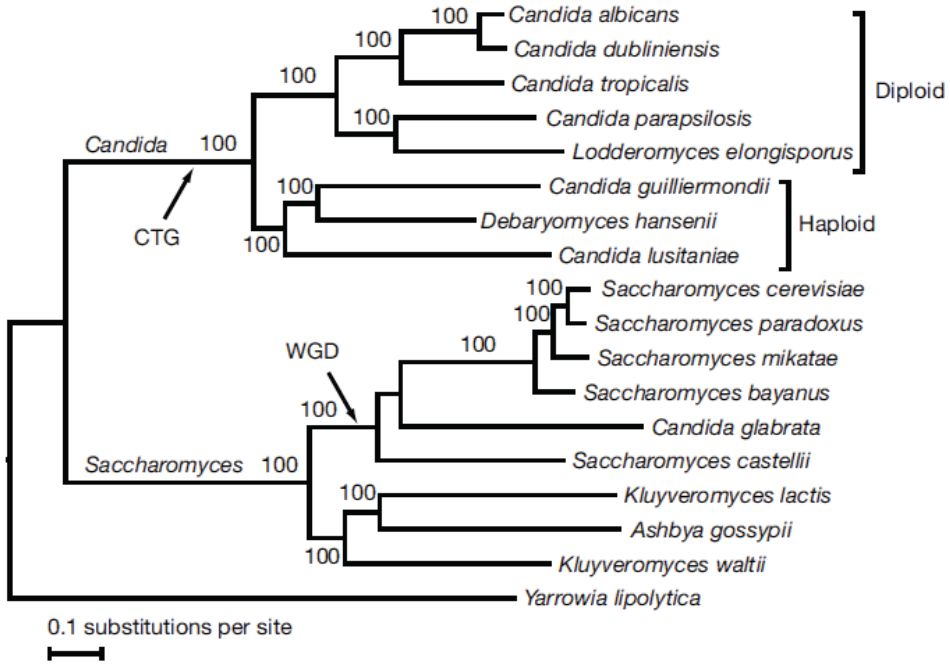


Figure 1-17. Phylogeny of sequenced *Candida* and *Saccharomyces* clade species. Adapted from Butler *et al.*, 2009.

1.4.3.1 $tRNA^{Ser}_{CAG}$, LeuRS and SerRS

The key player in CUG reassignment is a hybrid $tRNA^{Ser}_{CAG}$ (Figure 1-18) that contains identity elements for both SerRS and LeuRS, which permit its unusual dual recognition. Depending on whether the two synthetases recognize the $tRNA^{Ser}_{CAG}$, the charging with leucine and serine is distinct, hence the dissimilar reassignment status between CTG clade species (Santos *et al.*, 1993;Santos *et al.*, 1996;Suzuki *et al.*, 1997).

This novel $tRNA^{Ser}_{CAG}$ appeared through insertion of an adenosine in the middle position of the anticodon of a $tRNA^{Ser}_{CGA}$ (Massey *et al.*, 2003). This adenosine insertion displaced the splice-site of the anticodon loop intron by one nucleotide and transformed the 5'-CGA-3' anticodon into the 5'-CAG-3' anticodon, creating a serine tRNA that could decode the leucine CUG codon as serine (Suzuki *et al.*, 1997). The emergence of a 5'-CAG-3' anticodon in the body of a serine tRNA was problematic for efficient codon-anticodon interaction. Thus, this tRNA underwent a series of mutations that changed its decoding properties and, consequently, modulated CUG reassignment. Some of these mutations were the introduction of m^1G37 that prevents frameshifting and maintains decoding accuracy (Bjork *et al.*, 2001), and the mutation of U33 to G33 that distorted the anticodon arm. Such alterations resulted in an unique aminoacylation problem, due to the insertion of a major identity element for the LeuRS in a tRNA containing identity elements for the SerRS.

The main identity elements of the $tRNA^{Ser}_{CAG}$ for serylation are the discriminator base G73 (guanosine at position 73); the variable arm, which contains 3 conserved CG pairs that are directly recognized by the SerRS. A unique G33 (guanosine at position 33) induces a long-range distortion of the top of the anticodon stem of the $tRNA^{Ser}_{CAG}$ and lowers its leucylation efficiency (Perreau *et al.*, 1999), working as a leucylation anti-determinant. The $tRNA^{Ser}_{CAG}$ contains G73 in the discriminator position, which in *S. cerevisiae* functions as a negative identity determinant for leucylation, but it is still leucylated due to the presence of A35 (adenosine at position 35) and m^1G37 (modified guanosine at position 37) in its anticodon loop, which are directly recognized by the LeuRS (Soma *et al.*,

1996). Interestingly, within the CTG clade, different species have different $tRNA^{Ser}_{CAG}$ sequences which influence CUG decoding. For example, the $tRNA^{Ser}_{CAG}$ from *C. albicans* (Figure 1-18) possesses the G73, 3 GC base pairs in the variable arm, A35 and m^1G37. These features allow its recognition by both the SerRS and LeuRS and subsequent charging with serine and leucine. The two different aminoacyl-tRNAs (Ser-$tRNA^{Ser}_{CAG}$ and Leu-$tRNA^{Ser}_{CAG}$) then compete for CUG codons at the ribosome A-site because the mischarged Leu-$tRNA^{Ser}_{CAG}$ is not edited by the LeuRS or discriminated by the translation elongation factor 1 (eEF1A). In other words, leucine and serine are incorporated (3% and 97%, respectively) into the proteome at CUG positions *in vivo* (Gomes *et al.*, 2007). Conversely, in *C. cylindracea*, the CUG codon is decoded 100% as serine because the $tRNA^{Ser}_{CAG}$ has a A37 instead of m^1G37 and the LeuRS is not able to recognize A37 (Suzuki *et al.*, 1994;Suzuki *et al.*, 1997).

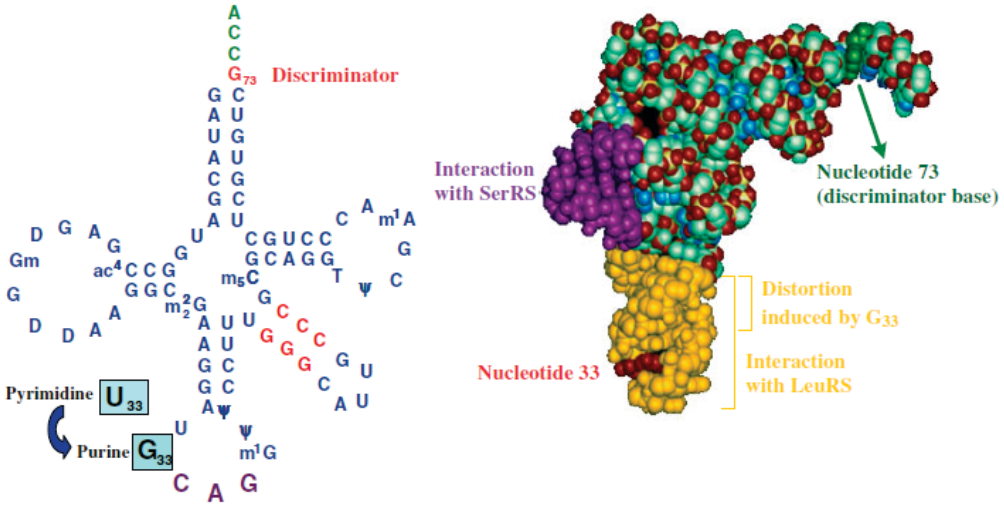


Figure 1-18. Secondary and tertiary structures of the *C. albicans* $tRNA^{Ser}_{CAG}$. The hybrid $tRNA^{Ser}_{CAG}$ molecule contains identity elements for both the SerRS, namely 3 GC base pairs in the extra arm (red) and the discriminator base (G73), and for the LeuRS the A35 and m^1G37 in the anticodon loop. The distortion indicated in the three-dimensional model has been determined *in vitro* in solution. Replacement of G33 with U33 increases leucylation efficiency of the $tRNA^{Ser}_{CAG}$ *in vitro*, showing that the main function of G33 is to keep CUG ambiguity at low level. Adapted from Miranda *et al.*, 2006.

Additional proof that the $tRNA^{Ser}_{CAG}$ is the crucial molecule in this reassignment comes from the study of the *C. albicans* leucyl-tRNA synthetase (CaLeuRS) and seryl-tRNA synthetase (CaSerRS). Several studies demonstrate that these synthetases have significant amino acid identity with homologous LeuRSs and SerRSs from other organisms, including *S. cerevisiae*. Indeed, they fully complement *S. cerevisiae* LeuRS and SerRS null strains, indicating that CaLeuRS and CaSerRS recognize and charge efficiently *S. cerevisiae* leucine tRNAs and serine tRNAs. This supports the hypothesis that the key element in CUG reassignment is the $tRNA^{Ser}_{CAG}$ rather than the LeuRS or the SerRS (O'Sullivan *et al.*, 2001a;O'Sullivan *et al.*, 2001b;O'Sullivan *et al.*, 2001c).

1.4.3.2 Pathway for the CUG reassignment in the fungal CTG clade

The reassignment of CUG codons in the fungal CTG clade, pictured in Figure 1-19, started 275 ± 25 million years (My) ago with the appearance of the above mentioned $tRNA^{Ser}_{CAG}$. This process was initiated before the divergence between the *Saccharomyces* and *Candida* genera (170 ± 27 million years ago) (Massey *et al.*, 2003). Hence, for approximately 100 My, the atypical $tRNA^{Ser}_{CAG}$ competed with the existing $tRNA^{Leu}_{CAG}$ for CUG decoding. This situation led to incorporation of both leucine and serine at CUG positions, i.e., CUG ambiguity. The hybrid $tRNA^{Ser}_{CAG}$ was lost in the lineage that originated *Saccharomyces* spp. (maintaining leucine CUG identity) and was preserved in the lineage that originated the CTG clade (Massey *et al.*, 2003).

CUG ambiguity should have caused major protein structural disruption due to major chemical differences between leucine and serine (leucine is hydrophobic and is located in the hydrophobic core of proteins; serine is polar and is located on the surface of proteins) (Tuite and Santos, 1996). Remarkably, comparative genomics studies showed that such codon ambiguity erased 98% of the 26000 - 30000 CUG codons that existed in approximately 50% of the CTG clade ancestor genes, likely to minimize the negative impact of CUG ambiguity on the proteome.

Therefore, CUG reassignment involved the disappearance of old leucine CUG codons, with the concurrent emergence of new CUG codons. In other words, "old" CUGs were mutated to the frequently used leucine UUA and UUG codons, while the ''new'' CUG codons present in extant species of the CTG clade evolved recently from codons coding for serine or amino acids with similar chemical properties. These observations indicate that more than half of the proteins of the fungal ancestor had their structure affected and that the CUGs of the CTG clade species are phylogenetically unrelated to the CUGs of the other fungal species (Massey *et al.*, 2003;Butler *et al.*, 2009).

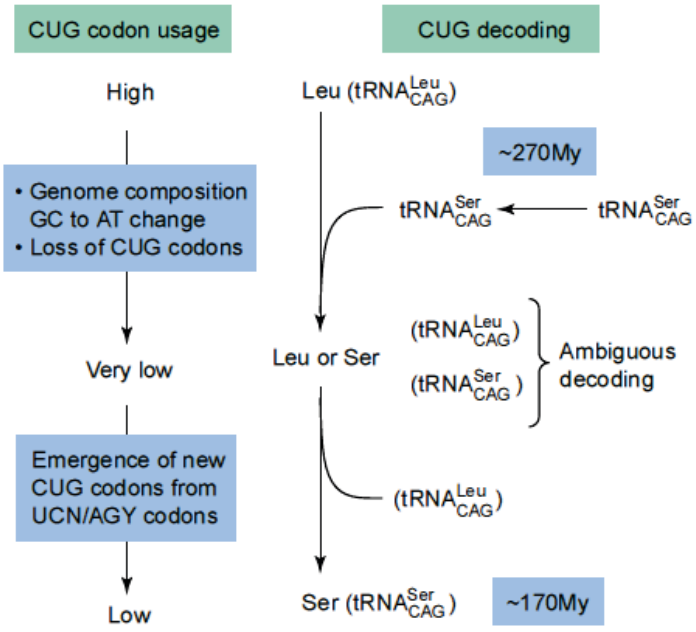


Figure 1-19. The evolutionary mechanism by which the CUG codon in *Candida* spp. has been reassigned from leucine to serine. Mutations in the anticodon of a tRNA$^{Ser}_{CGA}$ created a mutant serine tRNA with a 5'-CAG-3' anticodon (tRNA$^{Ser}_{CAG}$) that was able to compete with the natural tRNA$^{Leu}_{CAG}$ decoder for the CUG codon, thereby creating an ambiguous CUG codon. This enabled CUG reassignment through selection of the mutant Ser-tRNA and elimination of the cognate Leu-tRNA. CUG codon ambiguity imposed strong selection against old CUG codons, which mutated to UUG and UUA codons. This resulted in low CUG codon usage. However, presence of the mutant tRNA$^{Ser}_{CAG}$ introduced a new selective pressure that enabled re-introduction of CUG codons via mutation of serine UCN codons or other codons that code for related (i.e. conserved) amino acids (Santos *et al.*, 2004).

As mentioned above, different *Candida* species translate the CUG codon with distinct meanings. This dissimilar reassignment status raises the hypothesis that the pathway of CUG reassignment is still evolving in the ambiguous group, while the decoding of the CUG codon as serine only should represent the final stage of the reassignment pathway (Suzuki *et al.*, 1997). Support for this theory comes from evidences that ambiguous *Candida* species have low copy numbers of the $tRNA^{Ser}_{CAG}$ gene and the CUG codon is rarely used, which is consistent with the concept that low serine CUG usage is important to minimize the negative impact of ambiguity. In *C. cylindracea* where CUG is decoded as serine only, CUG is used at high level and the $tRNA^{Ser}_{CAG}$ gene copy number is high (Suzuki *et al.*, 1997). Therefore, it is likely that high CUG usage may have appeared late in the evolution of the genus *Candida*, when ambiguity no longer restricts CUG usage. Hence, *C. cylindracea* has high copy numbers of the $tRNA^{Ser}_{CAG}$ and uses CUG as a main serine codon (Suzuki *et al.*, 1997).

1.4.4 *C. albicans* as a model system to study the evolution of the genetic code

In an attempt to reconstruct the CTG clade genetic code change in *S. cerevisiae*, Santos and colleagues engineered CUG ambiguity in this yeast by introducing the *C. albicans* $tRNA^{Ser}_{CAG}$ into a *S. cerevisiae* strain while maintaining the $tRNA^{Leu}_{UAG}$ that decodes the CUG codon as leucine (Santos *et al.*, 1996;Santos *et al.*, 1999). Phenotypic characterization of this ambiguous recombinant strain revealed that CUG ambiguity increases tolerance to numerous stress agents like oxidants, heavy metals and drugs (Santos *et al.*, 1999). CUG mistranslation also produced important phenotypic heterogeneity concerning colony morphology, cell shape and size. Moreover, alterations on the gene expression profile were revealed, namely in categories corresponding to molecular chaperones, proteasome activity and carbohydrate metabolism (Silva *et al.*, 2007). Altogether, these results revealed that the growth rate disadvantage that resulted

from CUG ambiguity could be mitigated by increased adaptation potential to new ecological niches (Silva *et al.*, 2004). Another revealing result was obtained by Silva and colleagues, using the same ambiguous *S. cerevisiae* strain. CUG mistranslation increased ploidy and created aneuploid cell lines. It also blocked mating and sexual reproduction and produced polyploid asexual cell lineages. These data suggested that CUG ambiguity generates a genetic barrier that could have worked as a speciation mechanism for the CTG clade species (Silva *et al.*, 2007).

The study of CUG reassignment in *C. albicans* has also shown that CUG ambiguity is an important phenotypic diversity generator (Gomes *et al.*, 2007;Miranda *et al.*, 2007). Serine and leucine are incorporated into the proteome with efficiencies of 97% and 3%, respectively, but leucine misincorporation can be increased up to 28% without visible effects on growth rate (Gomes *et al.*, 2007). Indeed, hypermistranslator cells of *C. albicans* exhibited high population heterogeneity, including aerial hyphae, long filaments and high frequency of white-opaque switching. These cells also upregulated lipase and proteinase secretion and flocculated easily (Miranda *et al.*, 2007). Interestingly, morphological variation, yeast-hypha transition, proteinase and lipase secretion and adhesins, are important *Candida* spp. virulence traits (Calderone and Fonzi, 2001), suggesting that CUG ambiguity may be relevant in the process of infection.

In the case of *C. albicans*, a genome–wide analysis of CUGs helped understanding the global impact of CUG ambiguity. The *C. albicans* genome encodes 26,148 CUGs which are distributed over 66% of its genes. Considering the insertion of serine and leucine at each of these CUG positions (at a frequency of 1 to 38 CUGs per gene), *C. albicans* has the capacity to synthesize 283 billion different protein molecules from its 6438 genes. In other words, each *C. albicans* cell contains a unique combination of protein molecules and, consequently, the probability of finding two identical cells in a population is minute (Gomes *et al.*, 2007). Therefore, the *C. albicans* proteome is plastic, highly complex, and its size is unrelated to its gene pool. Whether novel functions are associated to mistranslated proteins remains to be elucidated (Santos *et al.*, 2011).

To get further insight on the impact of this nonconserved genetic code alteration in the proteome of *C. albicans*, Rocha and colleagues performed an extensive structural analysis of proteins containing CUG-encoded residues (Rocha *et al.*, 2011). A large fraction of proteins containing CUG encoded residues are involved in signaling pathways associated with morphological changes and pathogenesis. Data from 680 protein sequences were used and compared with orthologs from six related fungi. The analysis unveiled that codon reassignment led to a non-random genome-wide CUG redistribution that minimized protein misfolding events induced by the replacement of leucine with serine. Indeed, in *C. albicans* 90% of the CUG codons analysed are located in nonconserved positions where both leucine and serine can be introduced without major disruption of protein structure and function. This pattern was also preserved across the CTG clade species (Rocha *et al.*, 2011). This data was further reinforced with the crystal structures of the two isoforms of *Candida albicans* seryl-tRNA synthetase. Leucine or serine incorporation at the CUG position in *C. albicans* SerRS induced only local structural changes. In other words, *C. albicans* developed a highly complex mechanism to optimize protein structural robustness and produce functional plasticity through codon ambiguity (Rocha *et al.*, 2011).

1.4.4.1 *C. albicans* biology

Candida albicans is a usual resident of the gastrointestinal tract of humans and other warm-blooded animals, but is simultaneously the most common human fungal pathogen. It was first described in 1839 as a diploid organism with eight chromosomes that belongs to the *Saccharomycetaceae* family of ascomycete fungi (Odds *et al.*, 2007). The size of *C. albicans* genome is 14.3-14.4 Mb encoding 6,107-6,159 genes, depending on the strains (SC5314 and WO1) that have been sequenced (Braun *et al.*, 2005;Butler *et al.*, 2009). The genome of *C. albicans* is particularly flexible and allows the generation of genetic variability in the absence of sexual recombination. An outstanding example of this plasticity is

the appearance of an extra isochromosome 5 when cells are exposed to azole antifungals. This extra copy of the chromosome arm carries the *TAC1* gene that encodes the transcription factor that upregulates the expression of *MDR1*, a drug efflux pump. This mechanism allows cells to develop antifungal resistance (Selmecki *et al.*, 2006;Selmecki *et al.*, 2009). Therefore, genome malleability is used by *C. albicans* to increase fitness in numerous situations.

One of the most remarkable features of this organism is its ability to grow with three distinct morphologies: unicellular yeast cells that divide by budding; pseudohyphal cells, where buds elongate and do not separate from the mother cell, producing filaments of elongated buds with constrictions at the septal junctions; true hyphae that consist of chains of tube-like cells with no constrictions at the septal junctions (Figure 1-20). Hyphal growth is believed to be one of the most important virulence factors and it is promoted by a variety of environmental conditions, namely temperature (37°C), serum, neutral pH, high CO_2, growth in embedded conditions and N-acetylglucosamine (Sudbery *et al.*, 2004).

C. albicans cells are also of two types, namely “white” or “opaque”. Cells that form smooth, white and dome-shaped colonies are oval-shaped (white), while cells that form flatten and grey colonies are larger and have the surface covered in pimples (opaque). Colonies of opaque cells often switch back to white with a frequency of about 1:1,000 colonies. This phenomenon called phenotypic switching has been deeply characterized in *C. albicans* and is under epigenetic control (Slutsky *et al.*, 1987).

To date, no complete sexual cycle has been observed in *C. albicans* despite the demonstration that an elaborate mechanism for mating does exist. The white-opaque switching is an essential component of this mating mechanism (Bennett and Johnson, 2005). Most *C. albicans* isolates possess functional genes essential for mating in chromosome 5. Although, the majority of these isolates are heterozygous for a and α mating type loci (each contain two mating cistrons a1 and a2, and α1 and α2). The two heterozygous loci produce a a1/α2 heterodimer that represses genes required for mating. Also, for mating cells must be in the opaque phase because they only mate efficiently with cells of the opposite mating type in the opaque phase. Occasionally, strains become homozygous for a mating

type, likely by becoming haploid for chromosome 5 followed by duplication of the remaining monosomic chromosome. Those cells are then able to switch to the opaque phase where they can mate with cells of the opposite mating type which are also in the opaque phase (Tsong *et al.*, 2006). Oddly, white cells homozygous for mating type that are not competent for mating can also participate in this complex mechanism. When these cells are exposed to the mating pheromone of the opposite mating type they form a biofilm, which apparently allows the germ tubes of rare opaque cells of opposite mating type to orient their growth towards each other to enable mating (Tsong *et al.*, 2006;Daniels *et al.*, 2006). After germ tubes of opposite mating types meet, they fuse and karyogamy takes place (Lockhart *et al.*, 2003). The zygote formed by mating of two diploid cells is tetraploid. Meiosis has never been observed in *C. albicans,* suggesting that tetraploid cells regain the diploid state through chromosome loss during which recombination has been detected. Therefore, the process of mating in *C. albicans* leads to a parasexual reproductive cycle that provides a route for genetic material exchange between different strains (Kim and Sudbery, 2011).

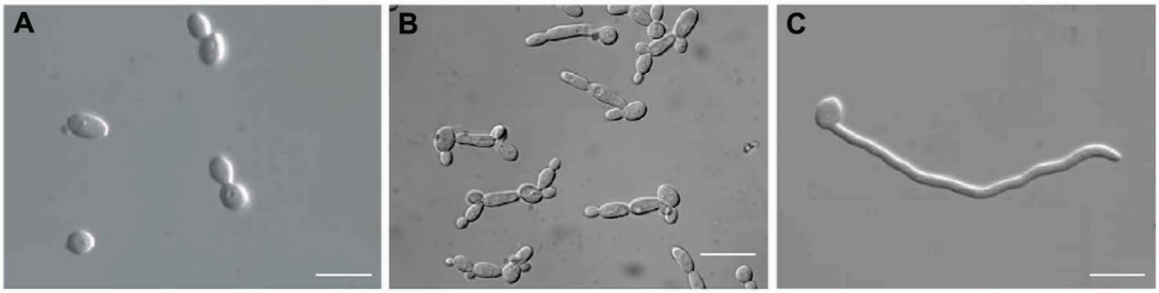


Figure 1-20. Different growth morphologies of *C. albicans.* (A) Differential interference contrast (DIC) image of cells with yeast morphologies. (B) DIC image of cells with pseudohyphae morphologies. (C) DIC image of cells with hyphae morphology. Bars represent 10 µm. Adapted from Sudbery *et al*., 2004.

1.5 Objectives of this study

The uniqueness of the CUG codon identity and previous data obtained in our laboratory strongly suggest that *Candida albicans* is an important model system to study the evolution of genetic code alterations. The objective of this thesis was to clarify the molecular mechanism of CUG reassignment and elucidate the functional role of CUG ambiguity. The specific objectives addressed in this PhD thesis were the following:

1. Produce new *C. albicans* strains with increased CUG ambiguity and revert CUG identity from serine back to leucine.

2. Carry out a detailed phenotypic characterization of the ambiguous and reverted strains.

3. Characterize the transcriptome of the above strains.

4. Sequence the genomes of those strains and determine the impact of CUG ambiguity on the genome.

2. Reversion of a genetic code alteration in the human pathogen *Candida albicans*

2.1 Abstract

The genetic code assigns codons to amino acids and defines how genes are translated into proteins. Despite this central role in cell function, its origin and evolution remain largely unknown and controversial and organisms with new codes have not yet been engineered. Several alterations to the standard code have been discovered over the last 40 years showing that the code evolves and is evolvable; however their biological relevance is poorly understood. *C. albicans*, the most prevalent human fungal pathogen, reassigned the CUG codon from leucine to serine and undergoes low levels of reassignment that increase under stress. Here, we constructed a series of strains that misincorporate leucine at serine CUG sites on a proteome wide scale at levels that range from 0.64% up to 100%, thus reverting CUG identity from serine back to leucine. Phenotype microarray analysis performed on the mistranslating set of strains revealed that they have few metabolic differences when compared to the naturally ambiguous strain. Still, CUG ambiguity generated high levels of phenotypic diversity and provided high adaptive potential in growth-based assays. This was most evident in strains with high levels of leucine misincorporation that exhibited increased tolerance to antifungals. This data highlights unanticipated roles of codon ambiguity on the evolution of phenotypic diversity and supports hypotheses postulating that ambiguous codons are re-assignable.

2.2 Introduction

The genetic code was established in the 1960s and its essential feature is the strict one-to-one correspondence between codons and amino acids (Crick, 1968). These codon-to-amino acid assignments are established by 20 aminoacyl-tRNA synthetases (AARSs) that recognize, activate and charge 20 proteinaceous amino acids onto tRNAs (Woese *et al.*, 2000;Ibba and Soll, 2004;Giege, 2006). Aminoacyl tRNAs (aa-tRNAs) are then transferred to the ribosome where their 3

letter anticodons read the 3 letter codons of messenger RNAs (mRNA) (Knight *et al.*, 2001). The code was originally designated as “universal” because of its extraordinary conservation in most organisms, from *E. coli* to humans. Also, it has been argued that the code must be “frozen” because codon reassignment would alter the amino acid sequences of numerous proteins and therefore have a lethal impact on the organism (or at least it would be strongly selected against) (Crick, 1968). This negative effect would be circumvented if the usage of the codon to be redefined could be lowered prior to reassignment (Osawa and Jukes, 1989), or if it is gradually adapted to the new assignment (Schultz and Yarus, 1996). However, in all situations accumulation of mutations in the genome is required, which is the major constraint “freezing” the genetic code. Despite this, the discovery of several codon reassignments in nuclear and organellar genomes from bacterial and eukaryotic species proved that the code is not “universal” or “frozen” and continues to evolve (Knight *et al.*, 2001;Santos *et al.*, 2004). The majority of known codon reassignments involve sense to nonsense codon changes (or *vice versa*) and only one nuclear sense-to-sense alteration is known so far, namely the reassignment of the CUG codon from leucine to serine in several *Candida* species (Santos *et al.*, 1997;Sugita and Nakase, 1999;Santos *et al.*, 2011).

Among code variants involving stop codons are the 4 glutamine codons of certain ciliates (Hanyu *et al.*, 1986), the 3 cysteine codons of other ciliates (Meyer *et al.*, 1991), and the tryptophan codon of *Mycoplasma* bacteria (Yamao *et al.*, 1985). Some of these reassignments involve codons whose identities change multiple times in close phylogenetic lineages and this suggests that certain taxonomic groups (e.g., the ciliates) are more prone to codon reassignment than others (van der Gulik and Hoff, 2011). Moreover, two non-canonical amino acids have also been naturally incorporated into the genetic code, namely selenocysteine in the active site of selenoproteins in a wide range of prokaryotes and eukaryotes (Allmang and Krol, 2006;Ambrogelly *et al.*, 2007;Yuan *et al.*, 2010) and pyrrolysine whose insertion occurs in the archeon *Methanosarcina barkeri* through reprogramming of the UAG stop codon (Krzycki, 2005;Gaston *et al.*, 2011b). Stop codons are more prone to reassignment than sense codons because they are less prevalent than sense codons (occurring only once per gene), their

reassignment causes minimal detrimental effects and also because release factors can mutate more freely (Knight *et al.*, 2001).

In recent years, remarkable efforts have been made to manipulate the genetic code in order to engineer new proteins. Non-canonical amino acids can be easily incorporated into proteins *in vivo* using orthogonal aminoacyl-tRNA synthetase/tRNA pairs together with nonsense codons. Up until now, more than 50 unnatural amino acids have been incorporated into proteins in *E. coli*, *S. cerevisiae*, *P. pastoris* and mammalian cell lines (Wang *et al.*, 2009;Young and Schultz, 2010;Hoesl and Budisa, 2012). Misincorporation of canonical amino acids is also being studied. Recently, UAG stop codons have been reassigned to glutamine (Gln) and tyrosine (Tyr) in a modified *E. coli* strain lacking both UAGs in essential genes and the release factor-1 gene (RF1) which decodes UAGs (Mukai *et al.*, 2010). Sense codons have also been reassigned to semi-conserved amino acids in *E. coli* through selective pressure incorporation (SPI) methodologies that activate amino acid misincorporation in quiescent cells to minimize toxic effects of codon ambiguity (Bacher and Ellington, 2001;Bacher *et al.*, 2005). Also, Turanov and colleagues (2009) showed that the genetic code of *Euplotes crassus* supports the insertion of two amino acids (selenocysteine and cysteine) by one codon UGA and that the dual use of the codon can occur within the same gene (Turanov *et al.*, 2009). Therefore, the genetic code is flexible and evolvable through forced evolution strategies, but how natural variation in codon - amino acid assignments emerges and is selected and the biological consequences of engineering the genetic code are poorly understood.

One of the most dramatic natural codon reassignments occured in fungal species of the *Candida*, *Debaryomyces* and *Lodderomyces* genera (fungal CTG clade) where leucine CUG codons are translated as serine by a mutant serine tRNA ($tRNA^{Ser}_{CAG}$) (Santos *et al.*, 1993;Sugita and Nakase, 1999). Among these species, *Candida cylindracea* incorporates 100% of serine at CUGs while *C. albicans* and most other CTG-clade species incorporate both serine (~97%) and leucine (~3%) (Santos and Tuite, 1995;Suzuki *et al.*, 1997). Previous studies have demonstrated that leucine misincorporation at CUGs is modulated by environmental cues and can be increased up to 28% without visible effects on

fitness (Gomes *et al.*, 2007;Miranda *et al.*, 2007), suggesting that CUG coding is highly flexible. We take advantage here of this unexpected flexibility to clarify the biology of codon reassignment and explore whether codon ambiguity can provide an effective mechanism to alter the genetic code.

In the present study, we developed a fluorescent reporter system to monitor CUG ambiguity *in vivo* and we used transfer RNA (tRNA) engineering and forced evolution methodologies to revert the atypical identity of *C. albicans* CUG codons from serine back to leucine. We started by infiltrating 20.6% of leucine at CUG positions on a proteome wide scale and increased this misincorporation in a stepwise manner up to 100%, re-creating *C. albicans* clones with the standard genetic code. Results revealed that the *C. albicans* proteome is highly tolerant to this genetic code alteration and that such unique manipulation generates an impressive array of phenotypes that can be advantageous under certain physiological conditions. Therefore, the CUG codons of the fungal CTG clade species have two true identities and these fungi may use one or the other depending on environmental and physiological conditions. This first artificial alteration of the genetic code provides important new evidence for a pivotal role of codon ambiguity in the evolution of the genetic code. The data suggests that the role of genetic code ambiguity in codon reassignment can be harnessed for producing organisms with new genetic codes and tools for synthetic biology.

2.3 Methods

2.3.1 Strains and growth conditions

Escherichia coli strain JM109 (*recA1 SupE44 endA1 hsdR17 gyrA96 relA1 thi* Δ*[Lac-proAB] F'[traD36 proAB-lacI lacZ*Δ*M15*) was used as a host for all DNA manipulations. *E. coli* cells were grown at 37ºC in LB broth medium or in LB 2% agar (Formedium) supplemented with ampicillin 50 µg/ml (Sigma-Aldrich).

Candida albicans SN148 (arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm434/ura3Δ::imm434iro1Δ::imm434/iro1Δ::imm434) (Noble and Johnson, 2005) was grown at 30°C in YPD (2% glucose; 1% yeast extract, and 1% peptone). *Candida albicans* strains T0, T1 and T2 were grown in minimal medium lacking uridine (0.67% yeast nitrogen base without amino acids, 2% glucose, 2% agar and 100 µg/ml of the required amino acids). Heterozygous $tRNA^{Ser}_{CAG}$ knock-out strains (T0KO1, T1KO1 and T2KO1) were grown in minimal medium lacking uridine and arginine, while homozygous deletion cells (T2KO2) were grown in medium lacking uridine, arginine and histidine. Cells were grown aerobically at 30°C unless indicated otherwise. All *C. albicans* strains used in this study are listed in Table 2-1 below.

Table 2-1. *Candida albicans* strains used in this study.

Strain	Genotype	Relevant description
SN148	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm^{434}/ura3Δ::imm^{434} iro1Δ::imm^{434}/iro1Δ::imm^{434}	Noble *et al* (2005)
T0	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm^{434}/ura3Δ::imm^{434} iro1Δ::imm^{434}/iro1Δ::imm^{434} RPS1/rps1Δ::pUA709 (URA3)	Naturally ambiguous strain carrying *yE*GFP reporter
T1	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm^{434}/ura3Δ::imm^{434} iro1Δ::imm^{434}/iro1Δ::imm^{434} RPS1/rps1Δ::pUA702 (URA3, Sc tLCAG)	Carrying 1 copy of the $tRNA^{Leu}_{CAG}$ gene
T2	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm^{434}/ura3Δ::imm^{434} iro1Δ::imm^{434}/iro1Δ::imm^{434} RPS1/rps1Δ::pUA706 (URA3, Sc tLCAG, Sc tLCAG)	Carrying 2 copies of the $tRNA^{Leu}_{CAG}$ gene
T0KO1	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm^{434}/ura3Δ::imm^{434} iro1Δ::imm^{434}/iro1Δ::imm^{434} RPS1/rps1Δ::pUA709 (URA3) tSCAG/tscagΔ::ARG4	Same as T0 with deletion of 1 copy of the $tRNA^{Ser}_{CAG}$ gene
T1KO1	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm^{434}/ura3Δ::imm^{434} iro1Δ::imm^{434}/iro1Δ::imm^{434} RPS1/rps1Δ::pUA702 (URA3, Sc tLCAG) tSCAG/tscagΔ::ARG4	Same as T1 with deletion of 1 copy of the $tRNA^{Ser}_{CAG}$ gene
T2KO1	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm^{434}/ura3Δ::imm^{434} iro1Δ::imm^{434}/iro1Δ::imm^{434} RPS1/rps1Δ::pUA706 (URA3, Sc tLCAG, Sc tLCAG) tSCAG/tscagΔ::ARG4	Same as T2 with deletion of 1 copy of the $tRNA^{Ser}_{CAG}$ gene
T2KO2	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm^{434}/ura3Δ::imm^{434} iro1Δ::imm^{434}/iro1Δ::imm^{434} RPS1/rps1Δ::pUA706 (URA3, Sc tLCAG, Sc tLCAG) tscagΔ::ARG4/tscagΔ::HIS1	Same as T2 with deletion of both copies of the $tRNA^{Ser}_{CAG}$ gene

2.3.2 Plasmid construction

The *C. albicans* plasmids used in this study were based on the integrative vector CIp10 (Murad *et al*., 2000). For expression of one copy of the mutant *S. cerevisiae* $tRNA^{Leu}_{CAG}$ gene in *C. albicans* SN148, a DNA fragment containing the *S. cerevisiae* tRNA gene was cloned into *Sal*I cloning site of CIp10 plasmid, yielding plasmid pUA701 (map in annex A). The mutant tRNA gene was amplified from pUA15 (Miranda *et al.*, 2007) using standard PCR protocols. The second copy of the gene was cloned into *Cla*I cloning site of plasmid pUA701, yielding plasmid pUA705 (map in annex A).

The reporter system was constructed on the basis of the codon-optimized yeast enhanced green fluorescent protein gene (Cormack *et al.*, 1997) and was assembled into CIp10, pUA701 and pUA705 in three different versions. Initially, a DNA fragment containing the *Ca*ACT1 promoter region, the *yE*GFP gene and the *Sc*CYC1 terminator was amplified from pACT1-GFP (Barelle *et al.*, 2004) and inserted between the *Xba*I and *Bam*HI cloning sites. This originated plasmids $pUA702_{201\text{-}UUA}$, $pUA706_{201\text{-}UUA}$ and $pUA709_{201\text{-}UUA}$, respectively (map in annex A). Then, the *yE*GFP gene was altered by site-directed mutagenesis to change the codon UUA in position 201 to codon CUG, yielding plasmids $pUA702_{201\text{-}CUG}$, $pUA706_{201\text{-}CUG}$ and $pUA709_{201\text{-}CUG}$ (map in annex A). The *yE*GFP gene was then altered by site-directed mutagenesis to change the codon UUA in position 201 to codon UCU, yielding plasmids $pUA702_{201\text{-}UCU}$, $pUA706_{201\text{-}UCU}$ and $pUA709_{201\text{-}UCU}$ (map in annex A). All plasmids used for strain construction are listed in Table 2-2.

Table 2-2. Description of the plasmids constructed for this study.

Plasmid	Description
CIp10	*Candida* Integrating vector (Murad *et al* 2000)
pUA709$_{201\text{-}TTA}$	plasmid CIp10 containing the *yE*GFP reporter system cloned between *Xba*I *and Bam*HI restriction sites; yEGFP contains a wt TTA-leucine codon at position 201
pUA709$_{201\text{-}CTG}$	plasmid CIp10 containing the *yE*GFP reporter system cloned between *Xba*I *and Bam*HI restriction sites; yEGFP contains a CTG-ambiguous codon at position 201
pUA709$_{201\text{-}TCT}$	plasmid CIp10 containing the *yE*GFP reporter system cloned between *Xba*I *and Bam*HI restriction sites; yEGFP contains a TCT-serine codon at position 201
pUA701	plasmid CIp10 containing 1 copy of the mutant *S. cerevisiae* $tRNA^{Leu}_{CAG}$ gene cloned into *Sal*I restriction site
pUA702$_{201\text{-}TTA}$	plasmid derived from pUA701 containing the *yE*GFP reporter system cloned between *Xba*I and *Bam*HI; *yE*GFP contains a wt TTA-leucine codon at position 201
pUA702$_{201\text{-}CTG}$	plasmid derived from pUA701 containing the *yE*GFP reporter system cloned between *Xba*I and *Bam*HI; *yE*GFP contains a CTG-ambiguous codon at position 201
pUA702$_{201\text{-}TCT}$	plasmid derived from pUA701 containing the *yE*GFP reporter system cloned between *Xba*I and *Bam*HI; *yE*GFP contains a TCT-serine codon at position 201
pUA705	plasmid CIp10 containing 2 copies of the mutant *S. cerevisiae* $tRNA^{Leu}_{CAG}$ gene cloned into *Sal*I and *Cla*I restriction sites
pUA706$_{201\text{-}TTA}$	plasmid derived from pUA705 containing the *yE*GFP reporter system cloned between *Xba*I and *Bam*HI; *yE*GFP contains a wt TTA-leucine codon at position 201
pUA706$_{201\text{-}CTG}$	plasmid derived from pUA705 containing the *yE*GFP reporter system cloned between *Xba*I and *Bam*HI; *yE*GFP contains a CTG-ambiguous codon at position 201
pUA706$_{201\text{-}TCT}$	plasmid derived from pUA705 containing the *yE*GFP reporter system cloned between *Xba*I and *Bam*HI; *yE*GFP contains a TCT-serine codon at position 201

2.3.3 Strains construction

2.3.3.1 Construction of T0, T1 and T2

In order to construct *C. albicans* T0, T1 and T2 strains (Table 2-1), integration of different plasmids were targeted to the RPS10 locus of the parental SN148 strain:

- pUA709 plasmid, bearing the fluorescent reporter system, was digested with *Stu*I and integrated at the RPS10 locus in SN148 strain to generate strain T0;
- pUA702 plasmid, bearing one copy of the mutant $tRNA^{Leu}_{CAG}$ gene, was digested with *Stu*I and integrated at the RPS10 locus of the parental SN148 strain to generate strain T1;
- pUA706 plasmid, bearing two copies of the mutant $tRNA^{Leu}_{CAG}$ gene, was digested with *Stu*I and integrated at the RPS10 locus of the parental SN148 strain to generate strain T2.

All *C. albicans* SN148 transformations with the respective plasmid were performed according to Walther and Wendland (2003) and transformants were selected in minimal medium without uridine (Walther and Wendland, 2003). The correct insertion of pUA709/ pUA702/ pUA706 was confirmed by PCR and DNA sequencing.

2.3.3.2 Construction of T0KO1, T1KO1 and T2KO1

In order to generate *C. albicans* heterozygous tSCAG/tscagΔ strains, disruption of one copy of the endogenous $tRNA^{Ser}_{CAG}$ was made using an ARG4 auxotrophic marker cassette targeted with 500nt-flanking homology for homologous recombination as described by Noble and Johnson (2005). However,

instead of using a fusion PCR procedure, the disruption fragment was assembled in an empty plasmid using regular cloning techniques.

$tRNA^{Ser}_{CAG}$ (*tScag*) 5' and 3' flanking sequences were amplified by PCR with primer pairs 1-2 and 3-4, respectively. Primers 5-6 were used to amplify the ARG4 auxotrophic marker (Figure 2-1 and Table 2-3). The flanking sequences and the selectable marker were then cloned into pUA515 (annex B) for the complete assembly of the disruption fragment. Primers 1-4 were used to amplify the *tScag::ARG4* fragment that was subsequently employed in different transformations:

- Strain T0 was transformed with the *tScag::ARG4* cassette originating strains T0KO1;
- Strain T1 was transformed with the *tScag::ARG4* cassette originating strains T1KO1;
- Strain T2 was transformed with the *tScag::ARG4* cassette originating strains T2KO1.

Transformants were selected in minimal medium without uridine and arginine. Disruption of the first copy of the endogenous $tRNA^{Ser}_{CAG}$ was verified by PCR of the flanks surrounding the inserted marker. The absence of the $tRNA^{Ser}_{CAG}$ gene was also verified by northern blot.

Table 2-3. Primers used for $tRNA^{Ser}_{CAG}$ disruption.

Primer No.	Oligo Name	Sequence 5' - 3'
1	oUA1588	TTTGAGCTCCGAACCAGTGGCTGTTTGTTTCCAG
2	oUA1589	AAAACTAGTCCAGTTATAAGCTTAAATGTAGAG
3	oUA1620	TTTGCTAGCGGACAGACTATTCAAGACAGCACAC
4	oUA1604	TTTGTCGACGTCAACCTCTATATGTGCATCCTCG
5 (ARG4)	oUA1606	TTTACTAGTTTGATGCTCGAGGGGGATCCG
6 (ARG4)	oUA1607	TTTGCTAGCTATAGAAATGCTGGTTGGAATGC
5 (HIS1)	oUA1608	TTTACTAGTGATTGGTCTGGTGCTAGCAAT
6 (HIS1)	oUA1609	TTTGCTAGCCGGTACCTGGAGGATGAGGAG

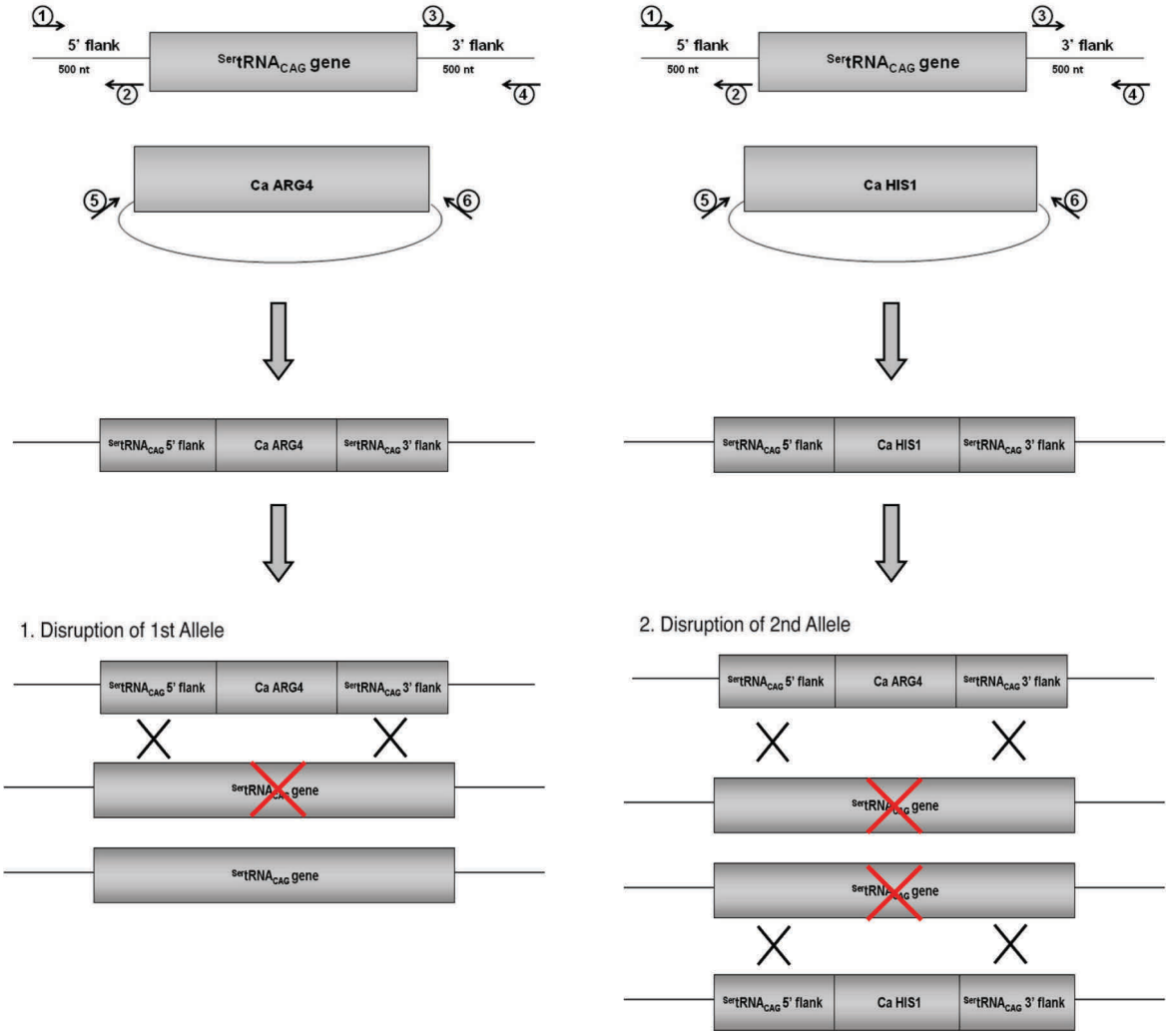


Figure 2-1. Deletion of *C. albicans* tSCAG (tRNA$^{Ser}_{CAG}$ gene). To delete tSCAG, disruption cassettes containing either the ARG4 or the HIS1 genes flanked by 500 bp of the 5` and 3` UTR of tSCAG were used. The disruption cassettes were generated by PCR, using the oligonucleotide primers 1-2, 3-4 and 5-6. ARG4 disruption cassette was transformed into *C. albicans* T0, T1 and T2 to generate strains T0KO1, T1KO1 and T2KO1, respectively. HIS1 disruption cassette was transformed into *C. albicans* T2KO1 to generate strain T2KO2, respectively.

2.3.3.3 Construction of T2KO2

In order to generate the *C. albicans* homozygous tscag∆/tscag∆ strain, disruption of the second copy of the endogenous $tRNA^{Ser}_{CAG}$ was made using an HIS1 auxotrophic marker cassette targeted with 500nt-flanking homology for homologous recombination as described by Noble and Johnson (2005).

$tRNA^{Ser}_{CAG}$ (*tScag*) 5' and 3' flanking sequences were amplified by PCR with primer pairs 1-2 and 3-4, respectively. Primers 5-6 were used to amplify the HIS1 auxotrophic marker (Figure 2-1 and Table 2-3). The flanking sequences and the selectable marker were then cloned into pUA514 (annex C) for the complete assembly of the disruption fragment. Primers 1-4 were used to amplify the *tScag::HIS1* fragment that was subsequently employed in transformation:

- The heterozygous strain T2KO1 was transformed with *tScag::HIS1* to produce the null mutant T2KO2.

Transformants were selected in minimal medium without uridine, arginine and histidine. Disruption of both copies of the endogenous $tRNA^{Ser}_{CAG}$ was verified by PCR of the flanks surrounding the introduced markers. The absence of the $tRNA^{Ser}_{CAG}$ gene was also verified by northern blot.

2.3.4 Yeast fitness assays

For the determination of growth rates, *C. albicans* strains were grown overnight at 30°C in liquid YPD medium as a pre-culture. For main cultures, fresh medium (YPD) was inoculated with cells from the pre-culture to an OD_{600} of 0.02. This culture was grown at 30°C, 180 rpm, until late stationary phase. Growth rate was monitored by the change in the natural logarithm of the optical density at 600 nm of the culture at the indicated time points. At least three independent experiments were used for each strain.

Cell viability was assessed using the colony forming units assay. Cell density in liquid culture was determined by counting the number of cells using a Neubauer chamber. Then the dilution factor necessary to dilute cells to 10^3 cells per ml was determined and cells were dispersed by vigorous vortexing and diluted in sterile water. 0.1 mL of diluted cells were plated onto an appropriate agar medium and incubated for three days at 30ºC. The colonies were counted and plating efficiency was determined. The procedure was repeated at the indicated time points and data are the result of three independent experiments with replicas.

2.3.5 Northern blot analysis

2.3.5.1 tRNA extraction

For total RNA extractions, 250 ml cultures grown overnight in minimal medium were harvested at an OD_{600} of 1.0-1.5 and the pellets were frozen overnight at -80ºC. Cells were resuspended in hot acid-phenol:chloroform 5:1 (pH 4.7) and TES buffer (10 mM Tris pH 7.5, 10 mM EDTA, 0.5% SDS). Cell suspension was incubated for 1 hour at 65ºC with repeated shaking every 10 minutes. The aqueous phase containing RNAs was separated from the phenolic phase by centrifugation at 8000 *g*, for 30 minutes, at 4ºC, and then transferred to a new tube and re-extracted with fresh chloroform:isoamyl alcohol 24:1. The aqueous phase was precipitated overnight, at -20ºC, with 3 volumes of cold absolute ethanol and 0.1 volumes of 3M sodium acetate. Tubes containing RNAs were harvested at 8000 *g*, for 30 minutes, at 4ºC, and pellets were resuspended in 0.1M sodium acetate pH 4.5. Total RNAs were applied to a 20 ml DEAE-cellulose column equilibrated with 0.1 M sodium acetate pH 4.5. tRNAs were eluted with 0.1 M sodium acetate/1 M sodium chloride and precipitated with 2.5 volumes of cold absolute ethanol, resuspended in 10 mM sodium acetate pH 5.0/1 mM EDTA, and stored at -80ºC (Santos *et al.*, 1993).

2.3.5.2 Northern blot

Fractionation of total tRNAs was carried out on 12-15% polyacrylamide (40% Acril:Bis) gels containing 8 M urea (0.8 mm thick, 30 cm long). In each gel slot, 50 µg of total RNA sample was loaded and gels were electrophoresed at 500 V for 16 hours. Fractioned tRNAs were localized by UV shadowing, the portion of the gel containing tRNAs was cut and transferred onto a nitrocellulose membrane (Hybond N, Amersham) using a Semy-dry Trans Blot (Bio-Rad). For hybridization, probes were prepared using 10 pmol of dephosphorylated oligonucleotide and 4 µl of ɣ-^{32}P-ATP (5000Ci/mmol) (Perkin Elmer) in 1x T4 kinase buffer, 10 mM spermidine and 16 units of T4 kinase (Takara). Labelling reaction was incubated at 37ºC for 1 hour and then the probe was extracted using phenol:chloroform:isoamyl alcohol (PCIA). The hybridization protocol was performed as described by Jacques Heitzler (Heitzler *et al.*, 1992). Membranes were pre-hybridized for 30 minutes at 53 ºC in a hybridization solution [5x Denhardt's solution (1% Ficol, 1% Polyvinylpyrrolidone and 1% Bovine serum albumin), 6x SSPE (3 M NaCl, 173 mM NaH_2PO_4, 25 mM EDTA) and 0.05% SDS]. Membrane hybridization was performed overnight in the above buffer using probes *GCGACACGAGCAGGGTTC* for detection of $tRNA^{Ser}_{CAG}$ and *GCGCCTCCGAAGAGATCA* for detection of mutant $tRNA^{Leu}_{CAG}$. Membranes were then washed 4 times (3 minutes each time) in 2x SSPE, 0.5% SDS at 53ºC. The membranes were exposed overnight with intensifying screens and developed using a Molecular Imager FX (Biorad).

2.3.6 Western blot analysis

2.3.6.1 Protein extraction and prepraration

Cells from 10 ml overnight cultures (OD 600nm between 0.5 and 0.8) were harvested by centrifugation, washed twice in 1x PBS (50 mM potassium phosphate, 150 mM NaCl pH 7.2) and resuspended in 0.3 ml of lysis buffer (50

mM PBS pH 7.0, 1 mM EDTA, 5% Glycerol, 1mM PMSF and EDTA-free protease inhibitors from Roche). The suspension was kept on ice, transferred to cryo-tubes and 1 volume of glass beads were added. The cell walls were digested using Precellys (5 cycles of 10 seconds in the beater and 2 minutes on ice). Then, tubes were centrifuged during 5 minutes at 13000 rpm and supernatants were removed into new tubes. This suspension was span down again for 10 minutes at 13000 rpm and a clear protein extract was obtained. Protein quantification was carried out using the Pierce BCA Protein Assay Kit. Aliquots of about 30 µg of extracted proteins were prepared in 15µL 1X SDS gel loading buffer (50mM Tris-HCl pH 6.8, 100mM DTT, 2% SDS, 0.1% bromophenol blue, 10% glycerol) and heated at 95ºC for 5 minutes, immediately before running the gel.

Samples were fractionated on a denaturing 15% SDS-polyacrylamide gel: water, acrylamide/bisacrylamide mix pH 7.0, Tris base, SDS, ammonium persulfate (APS) and TEMED. 15% resolving gels and 4% stacking gels were routinely prepared and after polymerization gels were mounted in an electrophoresis apparatus filled with running buffer (25mM Tris base, 250mM glycine pH 8.3, 0.1% SDS). 30 µg of each of the samples were loaded into the wells. The gels were run at 80 V until the dye front moved into the resolving gels; then the voltage was increased to 150V, until the dye reached the bottom of the gel (about 4 hours).

2.3.6.2 Western blot

After gel electrophoresis, samples were electroblotted onto nitrocellulose membranes (Hybond ECL, Amersham) using the Bio-Rad wet transferring system. One nitrocellulose membrane and six filters of the same size and cushions were placed in transfer buffer (25mM Tris base, 192mM glycine, 12% methanol). The transfer system was then mounted and transfer buffer was added to the blotting system. Blotting was allowed to run overnight at 4ºC. After blotting, membranes were peeled off from the gels and placed in 5% milk in TBS solution during 1 hour. The membranes were then washed in TBS-T (1X TBS, 0.1% Tween) and

incubated for 1 hour with an anti-GFP (FL) antibody (Santa Cruz) in an appropriate primary antibody solution in TBS-T. Then, membranes were washed 3 times with TBS-T, during 10 minutes, and incubated in the appropriate secondary antibody, namely an anti-rabbit IgG-peroxidase from goat (Sigma). Incubation was carried out during 1 hour in the dark. Finally, other 3 TBS-T washes were carried out in the dark for 10 minutes and signal detection was performed using the Odyssey Infrared Imaging System (Li-Cor Biosciences). A monoclonal anti-α-tubulin antibody (Sigma) was used as internal control for signal normalization purposes.

2.3.7 Epifluorescence microscopy

*yE*GFP expression was visualized in *C. albicans* cells by epifluorescence microscopy. Strains were grown overnight in liquid media at an OD_{600} of 2.0-2.5 and aliquots were spotted onto microscope slides. Fluorescence was detected using a Zeiss MC80 Axioplan 2 light microscope, equipped for epifuorescence microscopy with filter set HE38. Photographs were taken using an AxioCam HRc camera and images were analysed using ImageJ software. Mean fluorescence intensities (± standard deviation) were quantified in individual *C. albicans* cells expressing the reporter *yE*GFP Leu-UUA_{201} (positive control), Ser-UCU_{201} (negative control) and Ser/Leu-CUG_{201} (reporter). *yE*GFP fluorescence (intensity/pixel) was determined for at least 1000 cells in each case.

2.3.8 Phenotyping assays

The primary screen assayed the growth and colony morphology of two or more independent isolates of each strain. *C. albicans* strains were grown overnight at 30ºC in MM-uri and were then serially diluted to 1000 cells per ml. Approximately 50 cells were plated onto fresh agar plates and then allowed to grow at 30ºC for 7 days in a humidified incubator to prevent drying of the agar

surface. Colonies were photographed using a dissecting microscope equipped with AxioCam HRc camera and Axio Vision Software from Zeiss.

For the phenotypic screens, a series of different conditions were tested according to a protocol established by Homann and colleagues (2009), with a few alterations. Phenotyping media included nutritional cues, temperature, antifungal drugs, and a variety of stress conditions (Table 2-4) (Homann *et al*., 2009). Standardization of the initial inoculum for each strain was done by growing cell cultures overnight in liquid medium at an OD_{600} of 0.5-1.0. Cell density in liquid culture was determined by counting the number of cells using a Neubauer chamber. Then the dilution factor necessary to dilute cells to 10^8 cells per ml was determined and cells were dispersed by vigorous vortexing and diluted in sterile water. 0.1mL of diluted cells were subjected to a series of 10 fold dilutions that were then transferred to a 96-well plate. This format allowed the plating of strain dilutions with a 96-pin bolt replicator (Caliper) to each of the assay plates. The plates were incubated and photographed over the course of a week using an AxioCam HRc camera and Axio Vision Software from Zeiss. All images were imported and processed using ImageJ software. Upon import, each spot of the assay ($A_{id\ stress}$) and control plates ($A_{id\ nostress}$, YPD or minimal medium growth at 30ºC) was measured and a growth score was calculated (corresponding to the ratio between the area of the spot in the stress condition and the area of the same spot in the control situation). The growth score average of all the spots corresponded to the score for that specific strain (GS_m). When compared to the control strain T0 (GS_c), the scoring system classified the strength of the reduction or enhancement of growth relatively to control.

- GS_m = Growth score mistranslating strain;
- GS_c = Growth score T0 control strain;
- A_{id} = Measured area for isolate “i” and spot dilution “d”;
- n_{id} = the set of all isolates in a given dilution for the strain under study;
- $GS_m > GS_c$ = Growth enhancement;
- $GS_m < GS_c$ = Growth reduction;

$$GS_m = \frac{1}{n_{id}} \sum_{n \in id} \left(\frac{[A_{id}]stress}{[A_{id}]nostress} \right)$$

Equation 2.1

$$GS_c = \frac{1}{n_{id}} \sum_{n \in id} \left(\frac{[A_{id}]stress}{[A_{id}]nostress} \right)$$

Equation 2.2

2.3.9 Growth competition assays

Competition growth assays were performed in triplicate with pairs of ambiguous clones (T2 or T2KO2) and control clones (T0). Ambiguous clones were tagged with GFP while the control T0 was not, which allowed strain differentiation by observation of cell fluorescence. Each competitor of the pair was inoculated at 10^6 CFU into 5 ml of YPD broth supplemented with a stressor. Cultures were incubated at 30°C with vigorous shaking. Daily transfers of 100 µl of resultant broth culture into fresh medium were performed. Relative survival of each competitor was calculated from the broth culture at 24, 48 and 96 hours. Five sets of experiments were performed (each in triplicate) in the following conditions:

- Strain T0 vs T2 in YPD with fluconazole 0.5 µg/ml;
- Strain T0 vs T2KO2 in YPD with SDS 0.04%.

Table 2-4. Growth conditions used for phenotypic characterization of *C. albicans* strains.

Assay	Base medium	Growth temperature
16ºC/ 30ºC/ 37ºC/ 42ºC	YPD/ MM	16ºC/ 30ºC/ 37ºC/ 42ºC
$CaCl_2$ 50 mM/ 300 mM	YPD	30ºC/ 37ºC
NaCl 1 M/ 1.3 M	YPD	30ºC/ 37ºC
C-absent	MM without glucose	30ºC
C-Galactose	MM with 2% galactose	30ºC
C-Glycerol	MM with 3% glycerol	30ºC
C-Ethanol	MM with 2% ethanol	30ºC
Caspofungin 0.2 µg/ml	YPD/ MM	30ºC
Itraconazole 1 µg/ml	YPD/ MM	30ºC
Fluconazole 0.5 µg/ml	YPD/ MM	30ºC
Calcofluor White 20 µM	YPD	30ºC
SDS 0.04%	YPD	30ºC
Sorbitol 1.5 M	YPD	30ºC
Caffeíne 15 mM	YPD	30ºC
EDTA 0.75 mM	YPD	30ºC
Menadione 80 µM/ 90 µM	YPD	30ºC
Hydrogen Peroxyde 4.5 mM/ 6.0 mM	YPD	30ºC
pH 4.0/ 8.6/10	YPD + buffered glycine	30ºC/ 37ºC
$CuSO_4$ 13 mM/ 15 mM	YPD	30ºC
LiCl 300 mM	YPD	30ºC
Guanidine HCl 3 mM/ 5 mM	YPD	30ºC
Urea 25 mM	YPD	30ºC
Ethanol 5%	YPD	25ºC/ 30ºC
N-absent	MM without (NH4)2SO4	30ºC
Spider medium	Spider medium	30ºC/ 37ºC
Phloxine B 6.0 µM	Lee's Medium pH 6.8	25ºC/ 30ºC
Lee's Medium	Lee's Medium pH 4.5 / 6.8	30ºC/ 37ºC

2.3.10 Phenotype microarray profiling (Biolog)

Five *C. albicans* strains (T0, T1, T1KO1, T2KO1, T2KO2) were tested on PM 96-well plates (PM1-10 and PM21-25). The complete list of compounds assayed by PM1-10 and PM21-25 can be obtained at http://www.biolog.com/pdf/PM1-PM10.pdf or in annex E. PM uses tetrazolium violet reduction as a reporter of active metabolism (Bochner *et al.*, 2001). The reduction of the dye causes the formation of a purple color that is recorded by a camera at defined time intervals, providing quantitative and kinetic information about the response of the cells in the PM plates. Strains were grown overnight at 30ºC on YPD and then cells were picked with a sterile cotton swab and suspended in 15 ml inoculation fluid (IF-0, Biolog). Cell density was adjusted to 81% transmittance (T) on a Biolog turbidimeter. PM plates were inoculated (100 ml per well) using cell suspensions complemented with 1% (v/v) Dye Mix E (Biolog). PM plates were incubated at 30ºC in an Omnilog Reader (Biolog) and monitored automatically every 15 min for color changes in the wells. Readings were recorded for 48 h, and data were analyzed with Omnilog-PM software (release OM_PM_109M) (Biolog), which generated a time-course curve for tetrazolium color formation. Data from Omnilog-PM software were filtered, using average height as a parameter and processed with Bionumerics software (Applied Math, Kortrijk, Belgium). An average height threshold of 100 arbitrary omnilog units (AOUs) was chosen to identify the carbon and nitrogen sources used by the strains (background curves in the control wells showed an average height of around 70 AOU). Two replicates (all trays) for each strain were performed (Bochner *et al.*, 2010).

2.4 Results

Gain of function GFP assays to quantify CUG ambiguity in *C. albicans*. In *Candida*, the reassignment of leucine CUG codons to serine evolved over 275 ± 25 million years through an ambiguous codon decoding mechanism (Santos *et al.*, 1996;Massey *et al.*, 2003). Ambiguity arose from competition of a mutant $tRNA^{Ser}_{CAG}$ with wild-type $tRNA^{Leu}_{CAG}$ for CUG decoding (Suzuki *et al.*, 1997;Massey *et al.*, 2003). Later, the disappearance of the wild type $tRNA^{Leu}_{CAG}$ from the *Candida* ancestor left the CUG decoding exclusively to the mutant $tRNA^{Ser}_{CAG}$. This tRNA could interact with seryl-tRNA synthetase (SerRS) and leucyl-tRNA synthetase (LeuRS) and maintained this ambiguous status on the majority of *Candida* species due to leucine mischarging of the hybrid $tRNA^{Ser}_{CAG}$ (Suzuki *et al.*, 1997).

Previous studies performed in our laboratory used direct protein mass spectrometry analysis to prove that the CUG codon is decoded as both serine and leucine *in vivo* in *C. albicans* (Gomes *et al*., 2007). Although useful, quantification of ambiguity using mass spectrometry methodologies can be complex and it does not allow for real-time quantification of misincorporation *in vivo* (for example in individual fungal cells during the infection process). To circumvent this limitation, we have developed a fluorescent reporter system to monitor ambiguous CUG decoding *in vivo* in *C. albicans* cells. The system was based on the yeast-enhanced green fluorescent protein (*yE*GFP) gene (Cormack *et al.*, 1997). This construction has the 239 amino acids encoded by optimal codons for *C. albicans* and GFP fluorescence is maximized by two mutations in the chromophore (Cormack *et al.*, 1997). *yE*GFP expression driven by the ACT1 promoter (pACT1-GFP) is widely used in *C. albicans* because expression levels are constant during growth in yeast and hyphal forms, and during growth *in vivo* in mouse models of systemic infection (Barelle *et al.*, 2004). Since the codon optimized *yE*GFP gene does not contain CUGs and has a leucine UUA codon at a critical position – position 201 – for fluorescence (Cormack *et al.*, 1997), we used this residue site to develop our CUG mistranslation reporter system (Figure 2-2). We started by producing a version of the *yE*GFP gene with a serine UCU codon at position 201

(Figure 2-2A). Expression of *yE*GFP$_{201\text{-}UCU}$ in *C. albicans* cells showed the expected loss of function caused by the presence of the non-synonymous codon. Interestingly, the mutant *yE*GFP could not be detected by western-blot (Figure 2-2B), indicating that serine-201 destabilizes *yE*GFP (Figure 2-2C). In contrast, reintroduction of the original leucine UUA codon in *yE*GFP produced functional *yE*GFP - protein could be detected (Figure 2-2B) and *C. albicans* cells were highly fluorescent (Figure 2-2C). On the basis of these results, we postulated that introducing the ambiguous CUG codon at this crucial position would result in functional protein only when leucine was incorporated (Figure 2-2A). This hypothesis was tested by expressing the three versions of the *yE*GFP (*yE*GFP$_{201\text{-}UUA}$, *yE*GFP$_{201\text{-}UCU}$ and *yE*GFP$_{201\text{-}CUG}$) in wild type SN148 *C. albicans* cells, using plasmid pUA709. Cells carrying these plasmids were grown overnight in liquid media at an OD_{600} of 2.0-2.5 and fluorescence was measured by epifluorescence microscopy. Photographs were taken and absolute GFP fluorescence intensities were quantified for individual yeast cells using ImageJ software. In each case, mean fluorescence intensities (± standard deviation) were calculated for at least 1000 individual cells in five different experiments (Figure 2-2D). Expression of *yE*GFP$_{201\text{-}UUA}$ was used as positive control (in this case, cells had the maximum level of fluorescence and incorporated 100% of leucine at position 201); the *yE*GFP$_{201\text{-}UCU}$ was the negative control (the level of autofluorescence of the cells corresponding to 0% of leucine incorporation); *yE*GFP$_{201\text{-}CUG}$ allowed for relative quantification of leucine and serine incorporation at CUG_{201}. Considering the absolute fluorescence of *yE*GFP$_{201\text{-}CUG}$ in *C. albicans* cells and after removing the *yE*GFP$_{201\text{-}UCU}$ autofluorescence, these cells had 1.45% ± 0.85 of fluorescence relative to the positive control *yE*GFP$_{201\text{-}UUA}$ (Figure 2-2D), indicating that the wild type strain incorporated 1.45% ± 0.85 of leucine at the CUG position. This value was in line with previous mass spectrometry data that showed that the natural level of ambiguity in *C. albicans* cells varies between 0.66% and 3.0% (Gomes *et al.*, 2007).

The new fluorescent mistranslation reporter system was further tested using *C. albicans* cells growing in different environmental conditions. Leucine misincorporation at the CUG codon was quantified in cells at 30ºC, 37ºC, pH 4.0

and 1.5 mmol/l H_2O_2, and levels of ambiguity ranged from 2.5% to 5.0% which matched the results previously obtained using the mass spectrometry reporter system (data not shown) (Gomes *et al.*, 2007). This confirmed that our methodology was suitable for precise quantification of mistranslation of the *C. albicans* serine CUG codon as leucine.

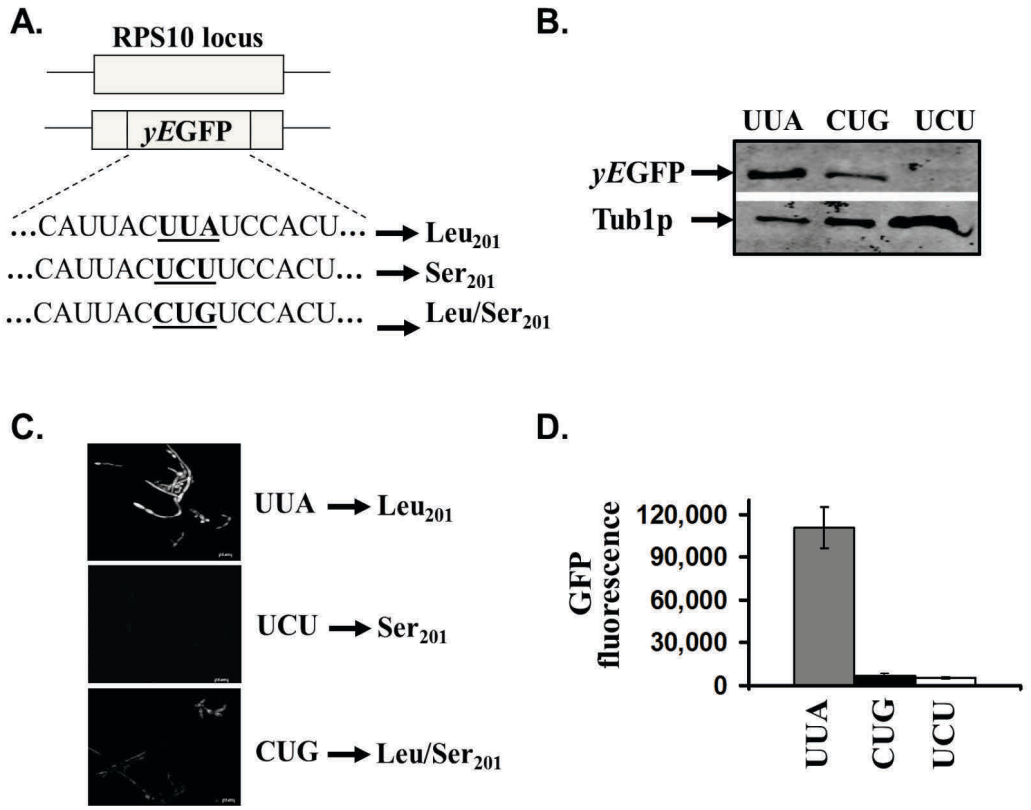


Figure 2-2. Fluorescent reporter system to quantify leucine insertion at CUG positions *in vivo*. A. The system was based on the yeast-enhanced green fluorescent protein (*yE*GFP) gene. The leucine UUA codon at position 201 was mutated to the ambiguous CUG codon and to a UCU serine codon. **B.** The different versions of the fluorescent reporter protein (27 kDa) were expressed in *C. albicans* and western blot analysis was performed on equal amounts of protein using an anti-GFP polyclonal antibody. No band was detected in the case of serine insertion at position 201 while leucine misincorporation provided a functional fluorescent protein. Tubulin (50 kDa) was used as an internal control. **C.** Cell fluorescence was determined using a microscope and images were processed using ImageJ software. The level of fluorescence was directly proportional to the amount of leucine incorporated at the CUG-201. **D.** Absolute quantification of fluorescence in SN148 *C. albicans* cells expressing the reporter system. Fluorescence of *yE*$GFP_{UUA\text{-}201}$ was the positive control and *yE*$GFP_{UCU\text{-}201}$ was the level of auto-fluorescence. *yE*$GFP_{CUG\text{-}201}$ fluorescence was calculated as 1.45% ± 0.85 of the maximum amount of fluorescence.

CUG codons of *C. albicans* have two true identities. The unexpected flexibility of CUG identity is emphasized by different levels of reassignment in diverse stress conditions and by the ability of cells to misincorporate up to 28% of leucine without a clear effect on fitness (Gomes *et al.*, 2007;Miranda *et al.*, 2007). This prompted us to push this ambiguity to its limit in order to reverse the atypical identity of *C. albicans* CUGs from serine back to leucine.

To accomplish this, we started by re-constructing in *C. albicans* the highly ambiguous status of CUG codons that existed 275 ± 25 million years ago in the CTG clade ancestor. For this, a *S. cerevisiae* $tRNA^{Leu}_{CAG}$ gene, derived from the *S. cerevisiae* $tRNA^{Leu}_{GAG}$ gene through mutation of the first anticodon wobble base (G to C), was used (Figure 2-3A). This mutant tRNA incorporates leucine at CUGs only (Miranda *et al.*, 2007). One copy of this mutant *S. cerevisiae* $tRNA^{Leu}_{CAG}$ gene and the fluorescent reporter system were co-inserted into plasmid pUA702 that was integrated into the RPS10 locus of *C. albicans* SN148. This manipulation generated recombinant clones (T1 clones) that harboured both the $tRNA^{Leu}_{CAG}$ gene and two copies of its $tRNA^{Ser}_{CAG}$ gene (Figure 2-3C). Competition between the heterologous $tRNA^{Leu}_{CAG}$ and the endogenous $tRNA^{Ser}_{CAG}$ for CUG codons was expected to increase CUG ambiguity. Since tRNA abundance is normally proportional to gene copy number (Percudani *et al.*, 1997;Goodenbour and Pan, 2006;Mahlab *et al.*, 2012), we also cloned a second copy of the $tRNA^{Leu}_{CAG}$ gene into plasmid pUA706 in order to generate T2 clones. These clones had two copies of the $tRNA^{Leu}_{CAG}$ gene and two copies of the $tRNA^{Ser}_{CAG}$ gene and therefore the respective tRNAs competed in equimolar levels for CUG codons at the ribosome A-site (Figure 2-3D). Finally, T0 clones bearing only the endogenous $tRNA^{Ser}_{CAG}$ gene and the reporter system were used as the wild type ambiguous control (Figure 2-3B). In all cases, the expression of the heterologous $tRNA^{Leu}_{CAG}$ gene was confirmed by northern blot analysis of tRNAs fractioned on 15% polyacrylamide gels. As expected, the T2 strain showed the highest level of $tRNA^{Leu}_{CAG}$ expression and T0 strain presented no expression of this leucine tRNA (Figure 2-6A). The level of ambiguity in T0, T1 and T2 strains was monitored using the gain of function fluorescent reporter system whose activity is dependent on incorporation of leucine at codon position 201. *C. albicans* T0 cells incorporated

1.45% ± 0.85 of leucine at that position and insertion of one copy of the *S. cerevisiae* $tRNA^{Leu}_{CAG}$ gene into the RPS10 locus (T1 clones) resulted in 20.61% ± 1.81 of incorporation. Integration of a second copy of the *S. cerevisiae* $tRNA^{Leu}_{CAG}$ (T2 clones) further increased leucine misincorporation from 20.61% to 67.29% ± 6.83 (Figure 2-7).

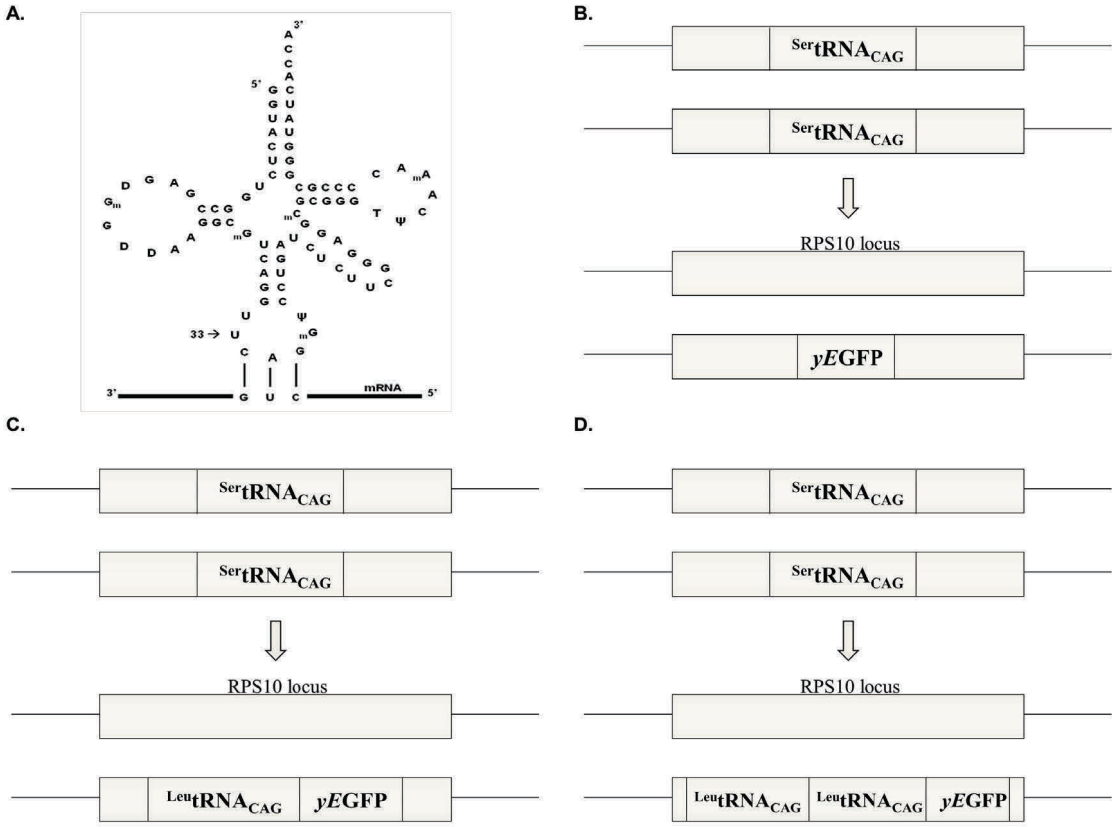


Figure 2-3. **Schematic representation of the stepwise construction of highly ambiguous *C. albicans* strains. A.** Engineered transfer RNA used to reconstruct the early stages of the CUG identity alteration in *C. albicans*. The anticodon of a *S. cerevisiae* leucine $tRNA_{GAG}$ was mutated to CAG for decoding CUG codons. **B.** In strain T0, plasmid pUA709 was used to insert the *yE*GFP reporter system into RPS10 locus of the *C. albicans* SN148 genome. **C.** In strain T1, plasmid pUA702 was used to insert the *yE*GFP reporter system and one copy of the $tRNA^{Leu}_{CAG}$ gene. This tRNA competed with two copies of the endogenous $tRNA^{Ser}_{CAG}$. **D.** In strain T2, plasmid pUA706 was used to insert the *yE*GFP reporter system and two copies of the $tRNA^{Leu}_{CAG}$ gene. Both copies of the mutant tRNA competed with two copies of the endogenous $tRNA^{Ser}_{CAG}$.

The second phase in this stepwise strain construction involved the deletion of one copy of the endogenous $tRNA^{Ser}_{CAG}$ gene (Figure 2-4A) to reduce the relative abundance of this tRNA. Heterozygous deletion strains were constructed from the previously built T0, T1 and T2 strains according to the protocol described by Noble and Johnson (2005) (see methods 2.3.3.2). Deletion of one copy of the $tRNA^{Ser}_{CAG}$ gene in T0 strain produced a strain (T0KO1) where CUG decoding was exclusively dependent on a single copy of the endogenous $tRNA^{Ser}_{CAG}$ gene (Figure 2-4B). Similarly, deletion of this wild type $tRNA^{Ser}_{CAG}$ gene in the T1 strain produced a strain (T1KO1) containing one copy of the $tRNA^{Leu}_{CAG}$ gene and one copy of the $tRNA^{Ser}_{CAG}$ gene (Figure 2-4C). Deletion of one copy of the $tRNA^{Ser}_{CAG}$ gene in the T2 strain produced a strain (T2KO1) bearing two copies of the mutant $tRNA^{Leu}_{CAG}$ gene and a single copy of the endogenous $tRNA^{Ser}_{CAG}$ gene (Figure 2-4D). Once again the fluorescent reporter system was used to measure the level of leucine incorporation in these clones. T1KO1 and T2KO1 cells showed an unexpectedly high incorporation of leucine (50.04% ± 1.64 and 80.84% ± 5.79, respectively), while T0KO1 cells showed the lowest level of mistranslation (0.64% ± 0.82) (Figure 2-7).

Finally, we attempted to revert CUG identity from serine back to leucine. To do so, we knocked out the second copy of the endogenous $tRNA^{Ser}_{CAG}$ gene in strain T2KO1. This generated T2KO2 strain which contains two copies of the $tRNA^{Leu}_{CAG}$ gene but lacked the $tRNA^{Ser}_{CAG}$ gene altogether (Figure 2-5B). Elimination of the endogenous tRNA was confirmed by northern blot analysis (Figure 2-6B). This strain misincorporated 98.46% ± 1.75 of leucine at the CUG position of the reporter, demonstrating that CUG identity was reverted from serine back to leucine (Figure 2-7). In order to determine whether the CUG codon could be translated by non-cognate *C. albicans* serine tRNAs, we have attempted to knockout the second copy of the $tRNA^{Ser}_{CAG}$ gene in T0K01 cells but we were unable to obtain viable clones (not shown). In other words, none of the $tRNA^{Ser}$ isoacceptors that decode the other 6 serine codons (UCN and AGR codons), neither the $tRNA^{Leu}_{IAG}$ that decodes leucine CUA, CUU and CUC codons were able to decode CUGs and compensate for elimination of the $tRNA^{Ser}_{CAG}$.

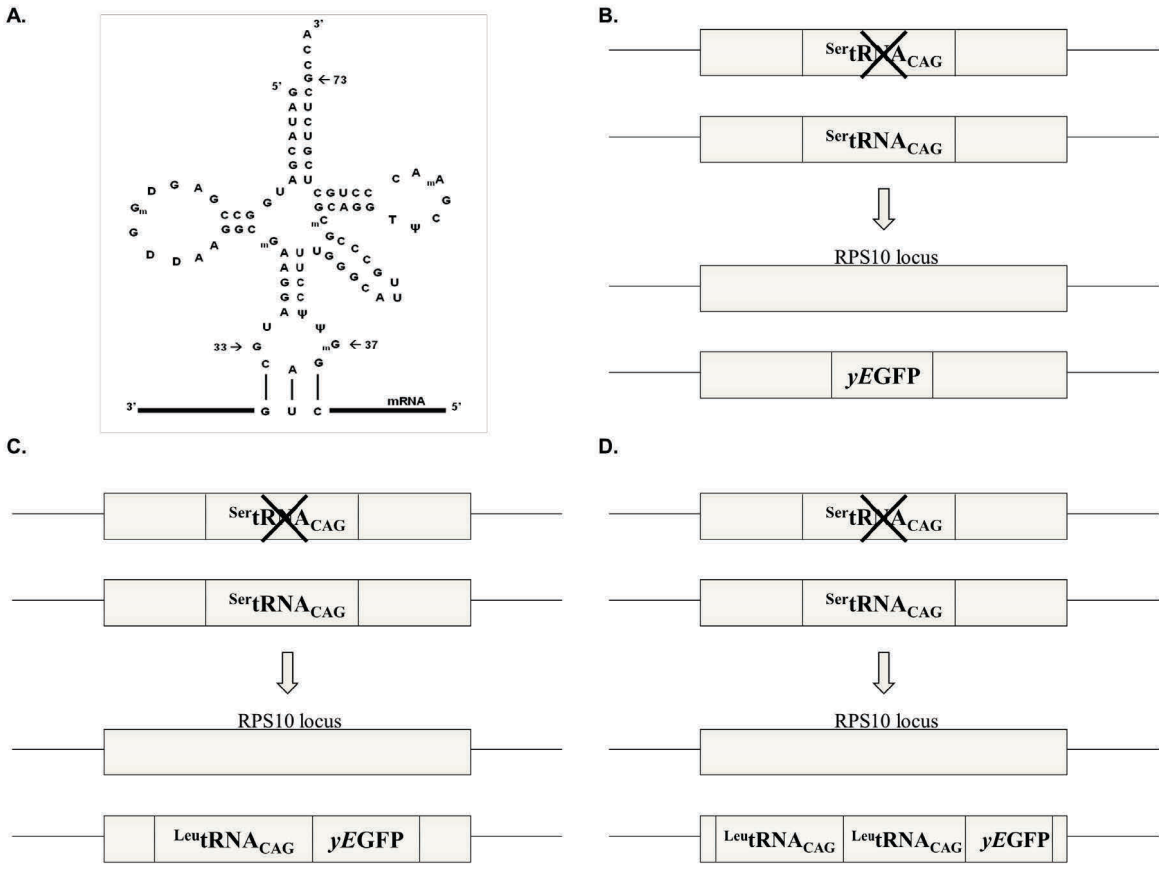


Figure 2-4. Schematic representation of the second phase of the construction of highly ambiguous *C. albicans* strains. A. One copy of the endogenous $tRNA^{Ser}_{CAG}$ gene was deleted in *C. albicans* T0 strain. **B.** Strain T0KO1 was produced from strain T0 by deleting one copy of the endogenous $tRNA^{Ser}_{CAG}$ gene. **C.** Strain T1KO1 was constructed from strain T1. The copy of the mutant $tRNA^{Leu}_{CAG}$ gene was kept and one copy of the endogenous $tRNA^{Ser}_{CAG}$ gene was deleted. **D.** Strain T2KO1 was produced from strain T2. Both copies of the mutant $tRNA^{Leu}_{CAG}$ gene were kept and one copy of the endogenous $tRNA^{Ser}_{CAG}$ gene was deleted.

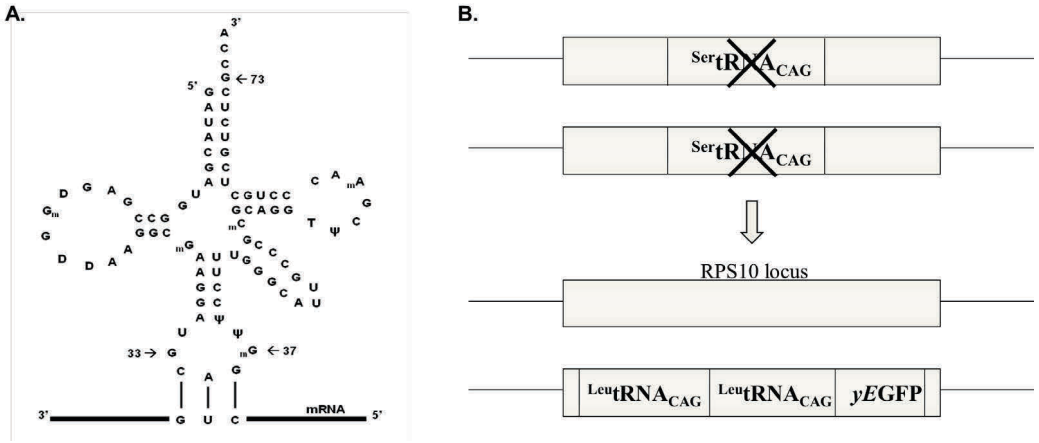


Figure 2-5. Schematic representation of the final phase of the construction of a *C. albicans* strain bearing the standard genetic code. A. Representation of the endogenous $tRNA^{Ser}_{CAG}$ gene that was knocked out in the *C. albicans* T2KO1 strain. **B.** Strain T2KO2 was built from strain T2KO1. The second copy of the endogenous $tRNA^{Ser}_{CAG}$ gene was deleted which left CUG decoding exclusively for the mutant *S. cerevisiae* $tRNA^{Ser}_{CAG}$.

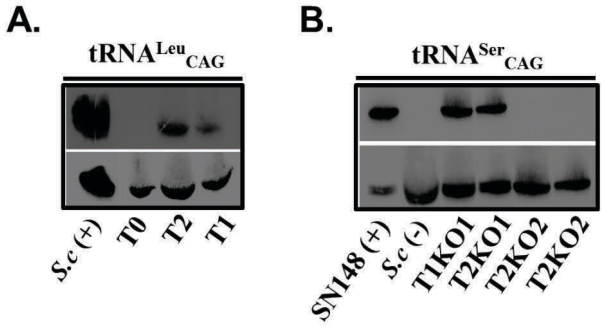


Figure 2-6. Monitoring tRNA expression by northern blot analysis. A. Strains with one and two copies of the mutant *S. cerevisiae* leucine $tRNA_{CAG}$ showed the expected expression in the northern blot. **B.** Strain T2KO2 showed no expression of the endogenous $tRNA^{Ser}_{CAG}$. $tRNA^{Thr}_{UGU}$ was used as an internal control for both assays (lower panel). S.c label indicates total tRNA extract from *S. cerevisiae*.

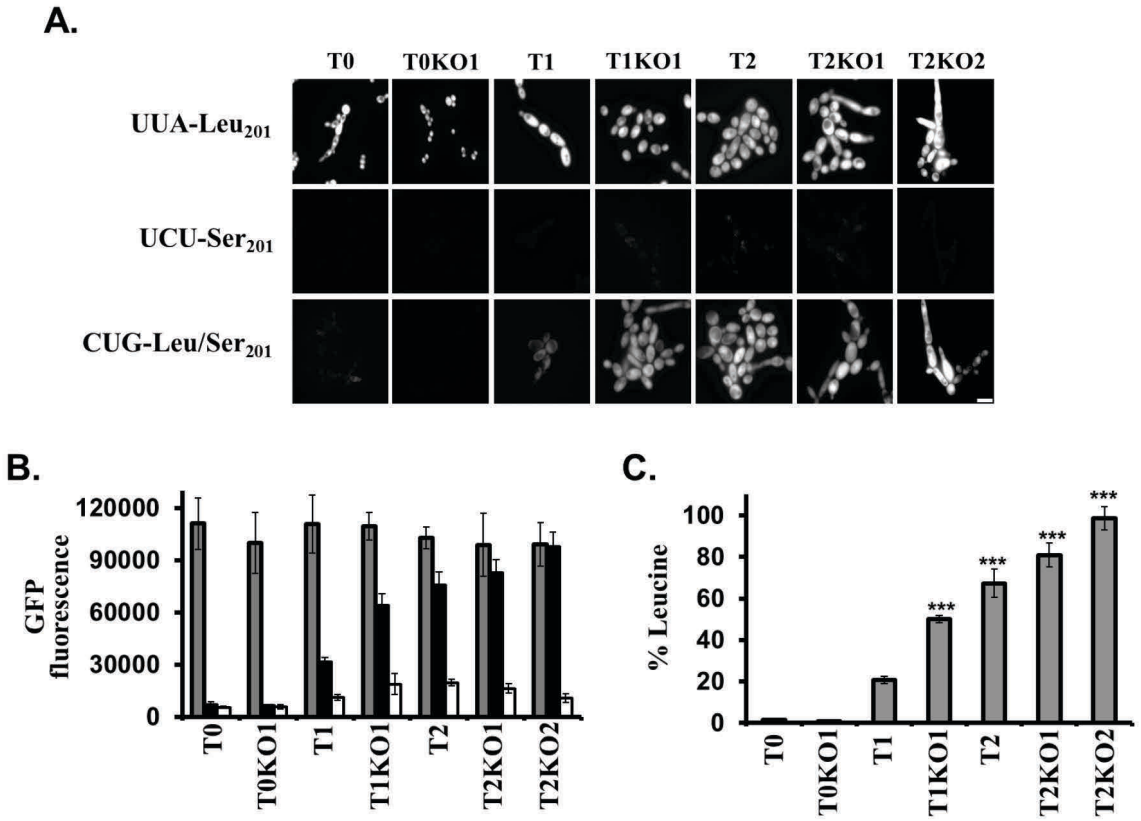


Figure 2-7. Quantification of leucine misincorporation in engineered *C. albicans* strains. A. Microscopic analysis of *C. albicans* cells expressing the three versions of the *yE*GFP reporter system. Incorporation of serine at this position inactivated *yE*GFP while leucine misincorporation provided a functional fluorescent protein. Scale bar represents 10 µm. **B.** Mean fluorescence intensities were quantified in individual *C. albicans* cells containing the positive control Leu-*yE*GFP$_{UUA-201}$ (grey bar), the negative control Ser-*yE*GFP$_{UCU-201}$ (white bar) and the reporter Ser/Leu-*yE*GFP$_{CUG-201}$ (black bar). GFP fluorescence (intensity/pixel) was determined for >1000 cells in each case. **C.** Leucine misincorporation at the CUG site was determined for each strain using normalized fluorescence data. In cells expressing 0, 1 and 2 copies of the mutant tRNA$^{Leu}_{CAG}$ gene, leucine incorporation was of 1.45±0.85%, 20.61±1.81% and 67.29±6.83%, respectively. Heterozygous tRNA$^{Ser}_{CAG}$ knock-out strains T0KO1, T1KO1 and T2KO1 had 0.64±0.82%, 50.04±1.64% and 80.84±5.79% of leucine incorporation. Homozygous deletion cells (T2KO2) showed 98.46±1.75% of leucine misincorporation at the CUG-201 position of *yE*GFP, indicating the complete reversion of CUG identity. Data represent the mean ± standard deviation of 5 independent experiments (***p<0.001 one-way Anova post Dunnett's comparison test with CI of 95%, relative to the T0 control cells).

Ambiguous CUG decoding generates phenotypic diversity. The stepwise genetic manipulation of *C. albicans* mentioned above originated 7 different strains, each of which possessing a different level of mistranslation (from 0.64% to 98.46% of leucine misincorporation). Transformation efficiency of these strains was significantly lower than that of T0 control clones (Figure 2-8C). However, the majority of clones that survived transformation showed a small decrease in growth rate. Indeed, T1, T2, T1KO1 and T2KO1 had approximately a 20% decrease in growth rate relative to T0. Remarkably, T0KO1 cells showed a 25% increase in growth rate relative to T0. Coincidently, these cells mistranslate at a lower level (0.64±0.82%) than the control cells which can explain the higher growth rate. The only strain that showed a significant decrease in growth rate was T2KO2 (60% decrease) (Figure 2-8A-B). Interestingly, clones with a standard genetic code (T2KO2) showed similar viability to that of the T0KO1 clones (both ends of the mistranslation spectrum) (Figure 2-8D).

To elucidate the extension of the impact of CUG ambiguity in *C. albicans* physiology, the engineered strains were studied in different growth conditions. The primary screen assayed growth and colony morphology of two or more independent isolates of each strain on solid media. In this case, strains that misincorporated more than 50% of leucine produced colonies with highly variable morphologies on agar plates, variable cell sizes and differentiation heterogeneity in liquid media (Figure 2-9). Colony and cell morphology variation increased with increasing levels of leucine misincorporation and the T2KO2 reverted strain produced morphotypes unrelated to those previously observed in the control strain, which produced *smooth* and *ring* colonies only. The *smooth* phenotype was characterized by dome-shaped colonies consisting of yeast cells and the *ring* phenotype had a wrinkled center with a smooth periphery. Clones T0KO1 and T1 produced a small proportion of *regular-wrinkled* colonies (Figure 2-9A), but strains that misincorporated more than 50% of leucine produced colonies with *irregular-wrinkled* and *jagged* morphologies (Figure 2-9B). These phenotypes were characterized by heavily rippled colonies composed of a mixture of yeast, pseudohyphal and hyphal cells (Figure 2-9B).

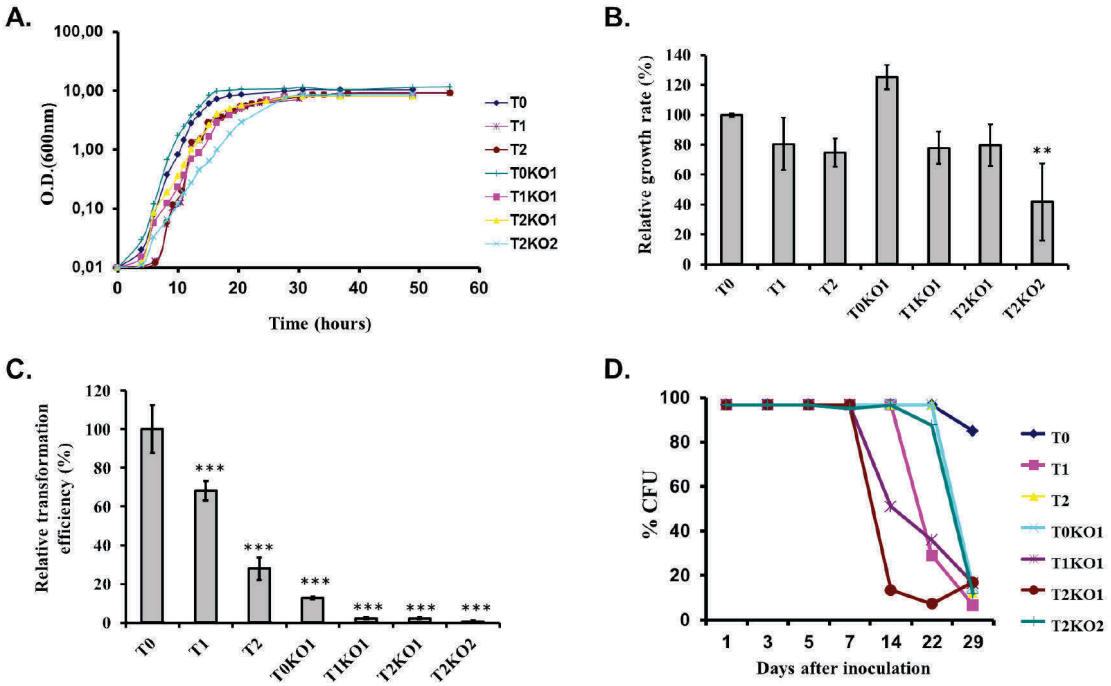


Figure 2-8. Growth rate, transformation efficiency and cell viability of the engineered *C. albicans* strains. A. Growth curves of the various *C. albicans* strains constructed on this study. Cultures were inoculated at an initial $O.D._{600}$ of 0.02 and were grown at 30º and 180 rpm until stationary phase. At the indicated time points the $O.D._{600}$ was measured. **B.** Relative growth rate of mistranslating strains was determined using exponential growth phase values relative to control T0 cells. Data represent the mean ± standard deviation of triplicates of 3 independent clones (**p<0.01 one-way Anova post Dunnett's comparison test with CI of 95%, relative to the T0 control cells). **C.** Relative transformation efficiencies of mistranslating *C. albicans* cells. Data represent the mean ± standard deviation of 3 independent experiments (***p<0.001 one-way Anova post Dunnett's comparison test with CI of 95%, relative to the T0 control cells). **D.** Viability of *C. albicans* cells was assessed using the colony forming units assay from 1 to 30 days in stationary phase. Results are presented in colony forming units (CFU) as a percentage of the cells plated.

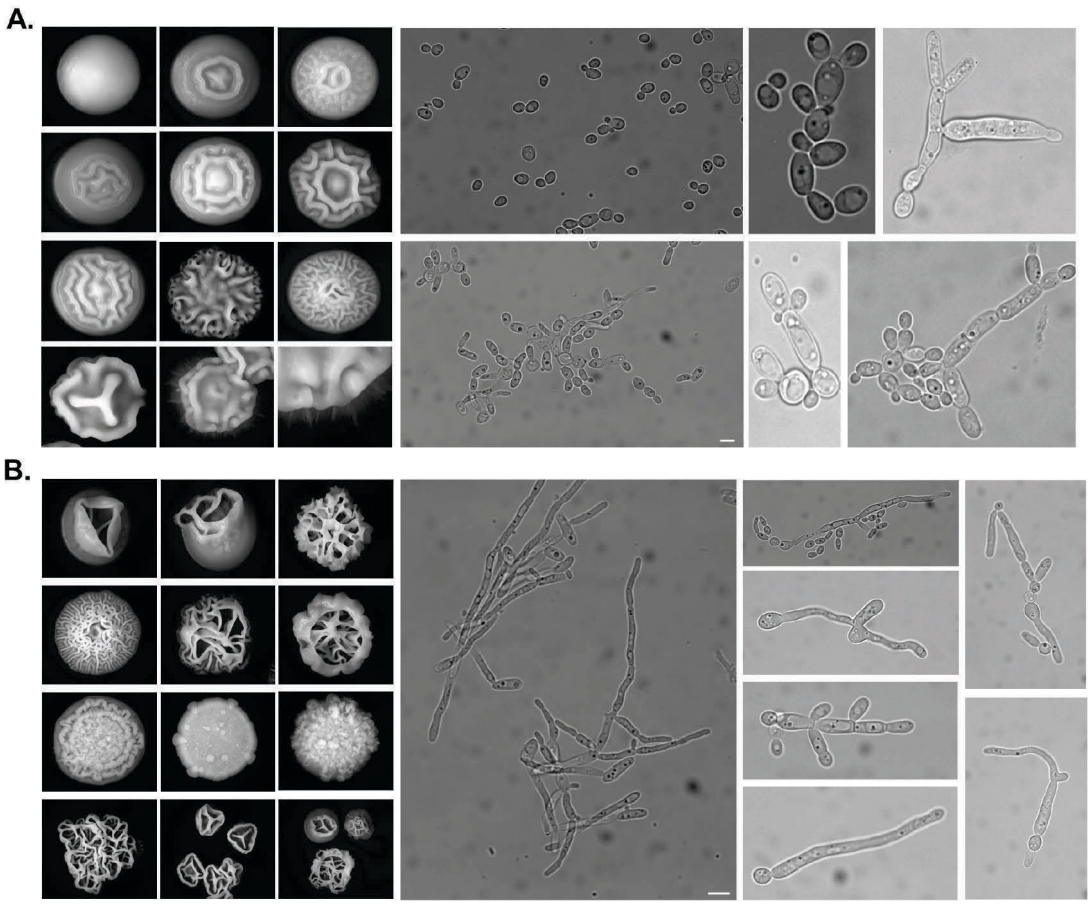


Figure 2-9. Colony and cell morphology phenotypes of ambiguous *C. albicans* strains. A. Repertoire of colony morphotypes ranging from smooth, ring, wrinkled and hyphae were observed. **B.** The morphological phenotype was exacerbated in knock-out mutants where irregular wrinkled and jagged colonies were observed. These knock-out mutants did not display the smooth or ring phenotypes and colonies consisted primarily of pseudohyphal and hyphal cell types. Colonies were p**h**otographed after 3 or 5 days of growth on YPD medium at 30ºC.

Colonies obtained in the primary screen were replated onto fresh agar plates to assess phenotypic stability. Switching among these different phenotypes is associated with the capacity of cells within colonies to undergo the yeast-to-hypha transition (Sudbery *et al.*, 2004). The strains with less than 50% of leucine misincorporation, namely T0, T0KO1 and T1 could switch from *smooth* to *ring* at high frequency, but the switch between *smooth* and *regular-wrinkled* was less frequent. The strains that misincorporated more than 50% of leucine (T1KO1, T1KO1 and T2KO2) switched frequently between *wrinkled* and *jagged* morphologies, but switching to the *smooth* phenotype was not observed. Therefore, increased CUG ambiguity increases morphological plasticity beyond the levels already described in the literature (Soll, 1992), indicating that *C. albicans* phenotypic variation is more extensive and flexible than previously reported.

To further test the metabolic flexibility and the response of all strains to environmental stress, they were incubated in a variety of different conditions on agar plates. These tests probed a broad spectrum of metabolic networks that included growth with different carbon sources, at different pH values (pH 4–10), at different temperatures (16°C–42°C), on agar plates supplemented with elevated cation concentrations or antifungal agents, or under other conditions that caused cellular stress (Table 2-4). When possible, drug/toxin/nutrient concentrations were calibrated in a way where both impairment and enhancement of growth relative to control could be observed. Independent isolates of each mistranslating strain were plated as a series of 10 fold dilutions on solid media using a 96-pin bolt replicator (Caliper). After 3-5 days of incubation, plates were photographed using Axio Vision Software from Zeiss. All images were imported and processed using ImageJ software and scored for growth phenotypes by comparison to the naturally ambiguous control strain (T0) included on the same plate. This approach generated individual growth scores for each mutant on each growth medium (annex D). Hence, the scoring system measured the reduction or enhancement of growth relative to control T0.

The phenotypic profiles produced by this methodology are provided in Figure 2-10A. The colour scale represents the range of phenotype variability from strong enhancement of growth (red) to strong reduction of growth (green). The intensity of the color represents the strength of the phenotype and black represents a phenotype that is indistinguishable from the T0 control strain. Although the changes in growth were idiosyncratic with each strain exhibiting a unique pattern of phenotypes, there were clear growth advantages under many conditions (Figure 2-10A and annex D). For example, all misincorporating strains grew faster than the control in presence of the oxidative stressors menadione and H_2O_2, in particular the strains T2, T2KO1 and T2KO2 (Figure 2-10A). Strains T0KO1 and T1 grew faster than the control in media supplemented with protein misfolding agents (urea and guanidine hydrochloride) (Figure 2-10A). However, all mistranslating strains grew slower than the control in the presence of salts, caffeíne and EDTA and also at 37ºC and 40ºC (Figure 2-10A). The most surprising results were the resistance of the T2K02 strain to SDS (Figure 2-10C) and $CuSO_4$ (Figure 2-10D), which could indicate that the cell wall has undergone major alterations in this strain. The behavior of the misincorporating strains in the presence of the main clinical antifungals, namely fluconazole, itraconazole and caspofungin is also noteworthy. Strains T2 and T2KO1 grew faster than the control strain in presence of fluconazole (Figure 2-10A) and also in presence of itraconazole, but grew poorly in presence of caspofungin, indicating that mistranslation may be linked with tolerance against azoles but not against echinocandins. Azoles are fungistatic heterocyclic synthetic compounds that inhibit the lanosterol 14α-demethylase, an enzyme that catalyzes a late step in ergosterol biosynthesis, and block demethylation of the C-14 of lanosterol, leading to depletion of ergosterol and accumulation of a toxic sterol intermediate in the plasma membrane (MacCallum *et al*., 2010).

Overall, there was variability between clones and environmental conditions, nevertheless, leucine misincorporating strains performed better than the control strain in numerous growth conditions indicating that despite the expected proteome disruption, the genetic code alteration could have positive outcomes.

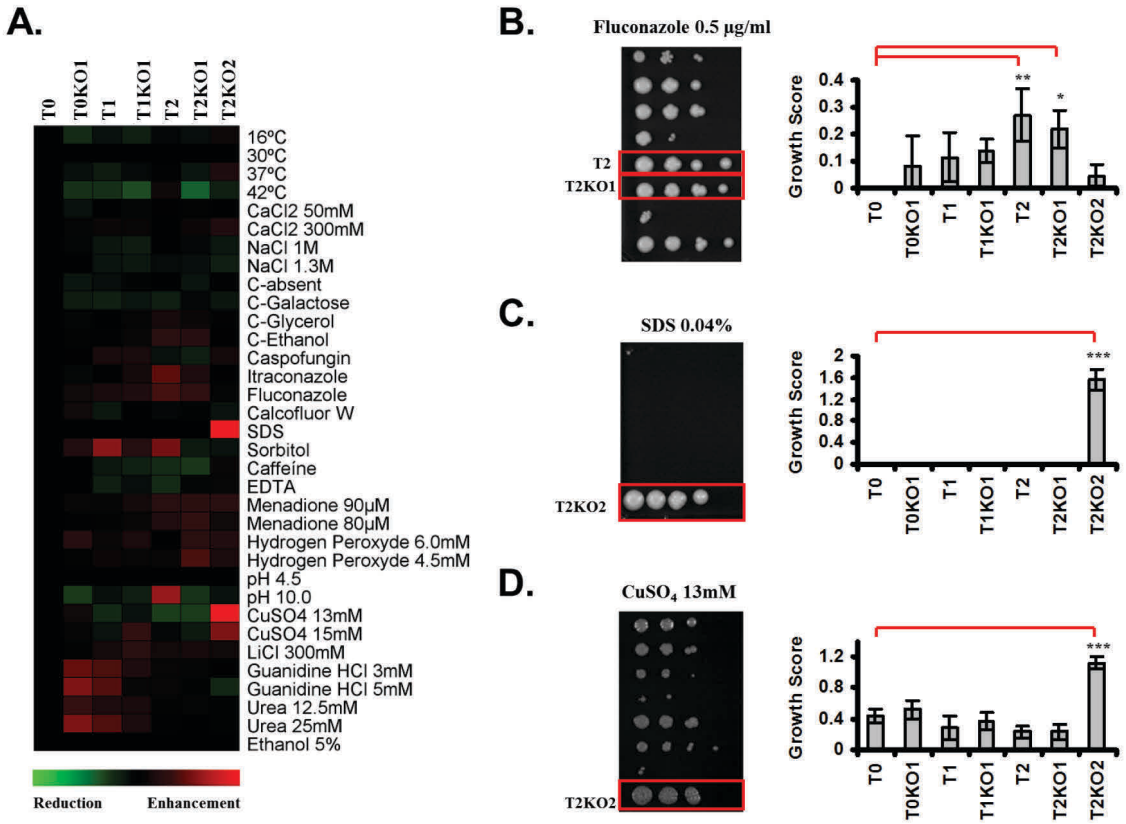


Figure 2-10. Phenotypic profiles of engineered *C. albicans* strains. A. Phenotypic screening of ambiguous *C. albicans* clones in basal, stress and specific nutritional media was carried out. Scores are represented by the color of the indicated square, and the intensity of the color represents the strength of the phenotype. Black squares represent a phenotype that is indistinguishable from the control strain T0. Green and red squares represent a reduction or enhancement, respectively. **B.** Growth of *C. albicans* strains in YPD medium supplemented with 0.5 µg/ml of fluconazole. Strains T2 and T2KO1 showed a significant growth advantage when compared to T0. **C.** Growth in YPD supplemented with 0.04% of SDS. Only the reverted strain T2KO2 showed a growth phenotype. **D.** Growth in YPD supplemented with 13 mM of $CuSO_4$. Only strain T2KO2 showed an advantageous growth phenotype. The growth score represents a ratio between growth in normal YPD medium and growth in YPD supplemented with a stress substance. Data represent the mean ± standard deviation of triplicates of 3 independent clones (***p<0.001; **p<0.01; *p<0.1 one-way Anova post Dunnett's comparison test with CI of 95%, relative to the T0 control cells).

To confirm the adaptive potential highlighted in the phenomics assay, strong phenotypes were selected for growth competition assays. Therefore, T2 and T2KO2 clones were paired with T0 and examined for competitive interaction in the presence of fluconazole and SDS, respectively. Equal numbers of colony forming units (CFU) of each competitor were inoculated into YPD broth supplemented with the corresponding stressor and incubated at 30°C with vigorous shaking. Surviving proportions of each competitor were calculated from the broth culture at 24, 48 and 96 hours, and strain differentiation was carried out by observation of cell fluorescence. This was possible because the mistranslating strains T2 and T2KO2 were tagged with GFP while the control T0 was not.

In broth culture containing fluconazole, the *C. albicans* T2 cells represented 60% of the mixed cell population after 24, 48 and 96 hours of growth (Figure 2-11A). The surviving proportion of mistranslating cells was further depicted in the case of T2KO2 cells growing in SDS (Figure 2-11B). After 48 hours of growth, the culture contained 0.04% of T0 cells and 99.96% of T2KO2 cells. After 96 hours of growth only T2KO2 cells were detected. These results strengthened the data obtained in the phenomics assay that suggested a positive outcome of the genetic code reassignment, namely by producing a growth advantage in several stress conditions.

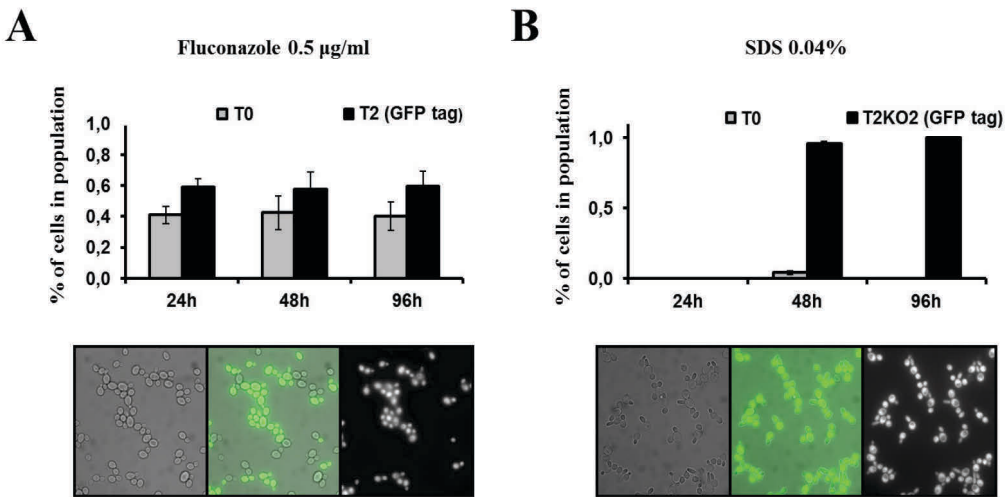


Figure 2-11. Growth competition assays between control T0 and mistranslating strains. Mistranslating strains T2 and T2KO2 were tagged with GFP and their behaviour was monitored for 96 hours under competitive interaction with control strain T0. **A.** When growing in presence of fluconazole, the T2 strain had a growing advantage. **B.** In the case of SDS, the advantage of the mistranslating strain T2KO2 (tagged with GFP) was clearer; after 96 hours of growth competition, T2KO2 cells represented 100% of the population. Five separate experiments were performed in triplicate.

Phenotype microarray profiling of ambiguous and reverted *C. albicans* strains. In a different approach, the mistranslating set of strains was tested using the Phenotype MicroArray (PM) platform developed by Bochner and colleagues (2001) and commercialized by Biolog Inc. The complete phenotype microarray for yeast cells included 15 pre-formulated 96-well microplates (PM1-10 and PM21-25 in annex E). Using these pre-formulated microplates, the protocol to profile cellular phenotypes in *C. albicans* involved a standard set of nearly 1500 assays for carbon (200 assays), nitrogen (400 assays), phosphorus and sulphur (100 assays) source utilization, nutrient stimulation (100 assays), pH and osmotic stresses (200 assays), and chemical sensitivities with 240 inhibitory chemicals (500 assays) (Bochner *et al.*, 2001;Bochner, 2009). Chemicals included in the array were selected based on their toxicity and interference with numerous cellular pathways, namely DNA replication, RNA transcription, protein synthesis, cell wall synthesis,

cell membrane synthesis and nutrient transport. Chemicals that are regularly found in natural environments were also incorporated in the array along with tests to measure the sensitivity of *C. albicans* strains to numerous inorganic chemicals, such as cations (Na^+, K^+, Fe^{3+}, Cu^{2+}, Co^{2+}, Zn^{2+} and Mn^{2+}) and anions (chloride, sulfate, chromate, phosphate, vanadate, nitrate, nitrite, selenite and tellurite) (Bochner *et al*., 2010).

The basis of the PM assay screen is a redox reaction that uses cell respiration as a reporter to profile growth phenotypes. The assay chemistry uses a tetrazolium dye (tetrazolium violet) to colorimetrically detect cell respiration. Reduction of this dye leads to the formation of a purple color and, as the dye reduction is irreversible, it accumulates in the well over a period of hours, amplifying the signal and integrating the level of respiration over time. The assay assumes that respiration is obligatorily coupled to cell growth and that, in a normal growth situation, cells must transport nutrients, catabolize and reform them, produce essential small molecule components, polymerize these into macromolecules to create and assemble subcellular structures. If all of these processes are working normally, the cell can grow and there will be a flow of electrons from the carbon source to NADH, down the electron transport chain of the cell, and, ultimately, onto the tetrazolium dye to produce the purple color. If one of these processes is working at an inferior rate it will restrict this flow and it will result in a decrease of the purple color. The degree of the restriction is replicated in the loss of purple color. The rate and magnitude of color formation in each well of the PM is monitored and recorded by the OMNILOG instrument (Bochner, 2003). The outputs of the OMNILOG are color-coded kinetic graphs. When two strains are compared, one is shown in red, another in green, and the overlap in yellow. Thus, yellow indicates no change in phenotype while red or green indicates a faster metabolism by the strain represented by that particular colour (Figure 2-13 B and C).

The PM assays were conducted in duplicate, and correlation plots from the independent runs indicated that replicate profiles were reproducible for plates PM1-2, PM8-9 and PM21-25 (Figure 2-12). On the other hand, assay plates PM3-8 did not pass reproducibility analysis. These assays tested nitrogen, phosphorus

and sulphur sources and several biosynthetic pathways and they proved to be very difficult to pass reproducibility due to the presence of high background in the chemical reaction. We have titrated amino acid supplements to single digit micromolar concentrations and still got overwhelmingly high background. This situation may be a consequence of the fact that all the ambiguous *C. albicans* strains have auxotrophies. According to several publications from the David Botstein's lab on starvation of yeast auxotrophs, they waste glucose upon amino acid starvation (Boer *et al*., 2008) which can cause the background dye reduction. This might be happening with our newly generated ambiguous strains causing the failure of the PM3-8 assays.

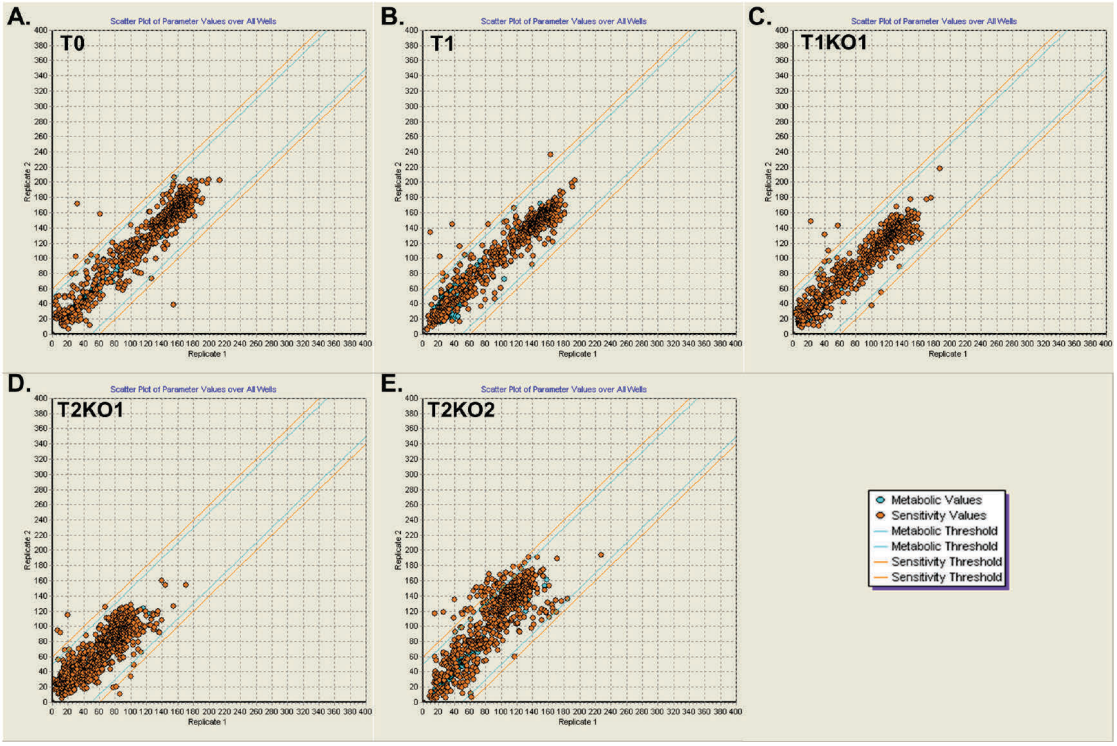


Figure 2-12. Reproducibility of PM tests. Average OminLog value was calculated over a 72-h time course for each replicate and plotted as replicate 1 (x-axis) versus replicate 2 (y-axis). Reproducibility analysis indicates the number of wells where the difference of average height between duplicate runs is above a threshold value, which is indicated in the graphs. Metabolic test values are represented by blue circles, and sensitivity test values are represented by yellow circles. PM3-8 were excluded from the analysis because of high background values. **A.** Reproducibility of PM screen in T0 strain. **B.** Reproducibility of PM screen in T1 strain. **C.** Reproducibility of PM screen in T1KO1 strain. **D.** Reproducibility of PM screen in T2KO1 strain. **E.** Reproducibility of PM screen in T2KO2 strain.

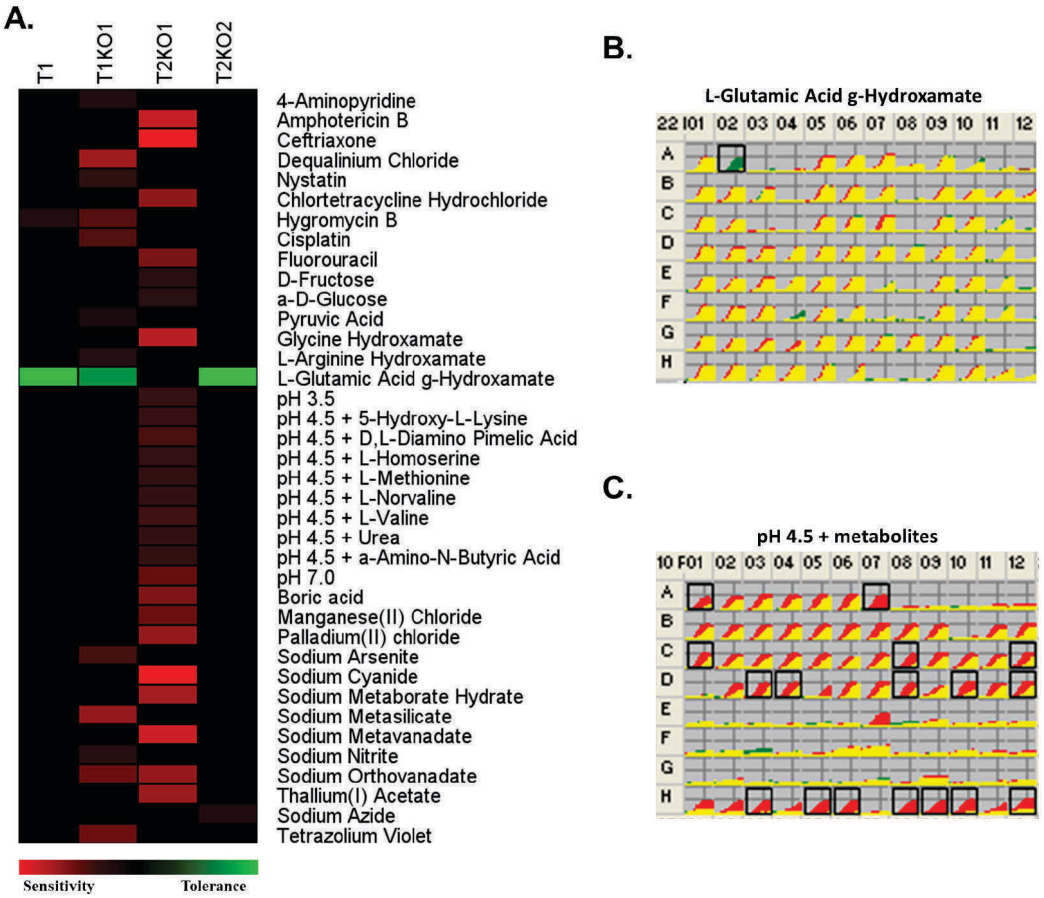


Figure 2-13. Phenomics of engineered strains. A. Black squares represent a phenotype that is indistinguishable from the control strain T0. Green and red squares represent gained/resistance and lost/sensitivity phenotypes of the test strain relative to the reference strain T0, respectively. The map is a summary of all phenotypes obtained for each strain in each assay. Data was obtained through the analysis of PM plates as in B and C. Duplicate arrays were run, and the average of the two runs was calculated. Only results that were significantly different in *both* runs are reported as significant. **B.** An example of the compared chemical sensitivities of two strains (T2KO2 vs T0). Assays were performed in Biolog PM22 microplates (annex E) that contained 24 assays, each in quadruplicate. Kinetic data were collected using the Biolog OMNILOG instrument and software. The curves show the time course (horizontal axis) of the amount of purple color formed from tetrazolium dye reduction (vertical axis) in each of the 96 wells. Data from T0 strain are shown in red, T2KO2 strain is in green, and yellow is the overlapping of the two kinetic curves. In well A2 (containing L-glutamic acid g-hydroxamate), strain T2KO2 had a higher metabolic rate which suggests a resistance phenotype relative to the reference strain. **C.** Assay performed in PM10 microplates (annex E) testing the response to variable pH (T2KO1 vs T0). Data from T0 strain are shown in red, T2KO1 strain is in green, and yellow is the overlapping of the two kinetic curves. In several wells of the plate, strain T0 had a metabolic advantage over T2KO1. This suggests a sensitivity phenotype of T2KO1 relative to the reference strain regarding pH 4.5 in the presence of several metabolites.

Analysis of PM1-2 (carbon sources), PM8-9 (osmolytes and pH) and PM21-25 (chemical sensitivity) was carried out using pairwise comparisons between the ambiguous strains and the control T0 strain. This profiling provided an overview of the main physiological characteristics of ambiguous and reverted strains (Figure 2-13A). As for T1 strain, only 2 phenotypes were observed: relative resistance to L-glutamic acid g-hydroxamate and relative sensitivity to hygromycin B (Figure 2-13A). These were common characteristics between T1 and T1KO1 that also exhibited resistance to L-glutamic acid g-hydroxamate and sensitivity to hygromycin B. Moreover, T1KO1 presented relative sensitivity to pyruvic acid, cisplatin, nystatin, tetrazolium violet, three toxic anions, sodium nitrite and L-arginine hydroxamate (Figure 2-13A). It is important to emphasize the sensitivity of these strains to several chemicals that are frequently used as antifungal drugs and whose mode of inhibitory action is distinct. Cisplatin works by binding to DNA, crosslinking two strands and preventing cell duplication (Siddik, 2003); nystatin acts by binding to sterols in the cell membrane with a subsequent change in membrane permeability allowing leakage of intracellular components (Khaled H. *et al.*, 2011); hygromycin B acts by inhibiting polypeptide synthesis (Borovinskaya *et al.*, 2008).

T2KO1 has more negative phenotypes than the others. We observed decreased metabolism for two carbon sources, glucose and fructose, and relative sensitivity at pH=3.5, pH=7 and pH=4.5 in the presence of several metabolites (Figure 2-13C). Furthermore, sensitivity was detected with amphotericin B, 5-fluorouracil, chlortetracycline, several toxic ions and glycine hydroxamate (Figure 2-13A). Once again, the mistranslating strain showed sensitivity to antifungal agents: amphotericin B increases the membrane permeability, 5-fluorouracil is a nucleic acid analogue and chlortetracycline blocks protein synthesis (Khaled H. *et al.*, 2011). Finally, the reverted strain T2KO2 had only two important phenotypes, namely relative resistance to L-glutamic acid g-hydroxamate (Figure 2-13B), and relative sensitivity to sodium azide (Figure 2-13A). The phenotype of resistance to L-glutamic acid g-hydroxamate was common between several mistranslating strains, namely T1, T1KO1 and T2KO2. Glutamic acid hydroxamate is a glutamic acid analogue that functions as a competitive inhibitor of Glu-tRNA synthetase,

therefore inhibiting protein synthesis. Its effectiveness in stopping growth depends on the concentration of glutamic acid hydroxamate and glutamate in the cell and the relative affinities of these two compounds for the activating enzyme. Therefore, if the Glu-tRNA synthetase was somehow altered during the reversion of the genetic code, the affinity of the glutamic acid hydroxamate for the enzyme may be changed which subsequently reduces the toxicity of this chemical. A similar mechanism of resistance was found in *E. coli* using a serine hydroxamate (Tosa and Pizer, 1971), where resistance to this amino acid analog is accompanied by changes in seryl-tRNA synthetase. The resistant *E. coli* strain possessed an enzyme with an altered structure which led to an increased K_i (Ser-Hdx) value and, in keeping with the competitive character of the inhibition, higher levels of serine hydroxamate were required to inhibit growth. These data confirmed that the activating enzyme is the point of inhibition and that the K_i (Ser-HDX) to K_m (Ser) ratio is critical in establishing the level of sensitivity (Tosa and Pizer, 1971). The same reasoning can be applied to explain the sensitivity of T1KO1 and T2KO1 to arginine and glycine hydroxamate, respectively.

Another clear result from the PM screen was the evident sensitivity of most mistranslating strains to several antifungals such as amphotericin B, ceftriaxone, nystatin, chlortetracycline, hygromycin B, cisplatin and fluorouracil. This was in marked contrast with the results from the growth-based phenomics assay where strains T2 and T2KO1 had a growth advantage when growing with fluconazole and itraconazole. Although, it is important to notice that these chemical agents belong to different classes of antifungals, hence strains must apply different mechanisms of resistance. For example, amphotericin B and nystatin are polyenes while fluconazole and itraconazole belong to the azole family of antimycotic agents (Ghannoum and Rice, 1999).

It is important to note that the PM screen uses tetrazolium redox chemistry in response to cell respiration as a reporter system which is considerably different from the previously described phenomics assay that measures growth on the agar surface. In fact, cell respiration can take place independently of cell growth and, occasionally, can measure phenotypes that do not lead to growth. Indeed, cells may respond metabolically by respiring but not growing. Furthermore, the PM

technology is limited because it uses minimal media. Minimal media is not optimal for strains that possess auxotrophies which may explain why many strains failed to grow at many of the PM conditions (PM3-8) and why few phenotypic differences were observed between control T0 and mistranslating clones. Also, production of NADH was measured for 48 hours while the growth-based assay was performed during 5 days. Generally, methods of comparison between strains and mutants must owe to differences in backgrounds, growth rate, and media. It should be noted that the nature of the Biolog plate contents makes precise determinations of proper inhibitory concentrations and media components difficult. Therefore, in some cases the concentrations of added nutrients may be too low or in a form that is unavailable to the target organism. This was not the case in the growth-based assay since drug/toxin/nutrient concentrations were calibrated in a way that both impairment and enhancement of growth relative to control could be observed. Despite the differences between the two assays, they both corroborate the variable phenotypic outcome of reverting the CUG identity in *C. albicans*.

2.5 Discussion

Natural codon reassignments highlight the existence of unexpected flexibility in the genetic code and suggest that the code is evolvable (Knight *et al.*, 2001). One example of such natural reassignments occurred in the genus *Candida* where thousands of nuclear leucine CUG codons were reassigned to serine (Santos and Tuite, 1995;Butler *et al.*, 2009). This reassignment was initiated 275 ± 25 million years ago by a mutant serine tRNA that acquired a 5'-CAG-3' leucine anticodon ($tRNA^{Ser}_{CAG}$) (Santos *et al.*, 1993;Santos *et al.*, 1996;Suzuki *et al.*, 1997;Massey *et al.*, 2003). During the early stages of the CUG reassignment, $tRNA^{Ser}_{CAG}$ competed with a $tRNA^{Leu}_{CAG}$ for CUG codons and consequently both leucine and serine were incorporated at CUG positions creating CUG ambiguity. For unclear reasons the $tRNA^{Leu}_{CAG}$ disappeared and the mutant $tRNA^{Ser}_{CAG}$ was selected leading to complete CUG reassignment (Santos *et al.*, 2004). Recently,

studies in *C. albicans* and *S. cerevisiae* provided fascinating insight on how the natural genetic code alteration may have evolved and on how it can produce physiological and phenotypic innovations through protein structural and functional diversification (Silva *et al.*, 2007;Gomes *et al.*, 2007;Miranda *et al.*, 2007;Moura *et al.*, 2009).

The majority of the species from the CTG clade, including *C. albicans*, incorporate both serine (~97%) and leucine (~3%) at CUGs due to the dual identity of the $tRNA^{Ser}_{CAG}$ (Santos and Tuite, 1995;Suzuki *et al.*, 1997). In 2007, our laboratory was able to prove that leucine misincorporation at CUG positions is modulated by environmental cues and this ambiguity can be increased up to 28% without visible effects on fitness (Gomes *et al.*, 2007;Miranda *et al.*, 2007). Considering this flexibility, we chose *C. albicans* as a working organism to try to achieve a complete replacement of codon identity and confirm whether codon ambiguity can provide an effective mechanism to alter the genetic code.

For CUG reassignment two critical steps were required, namely the expression of a mutant tRNA with the ability to decode CUG codons as leucine in *C. albicans* and the deletion of both copies of the endogenous $tRNA^{Ser}_{CAG}$. With this approach we were able to generate a series of strains with different levels of leucine misincorporation at CUGs, including one strain (T2KO2) with a standard genetic code. This reversion of identity of the CUG codon from serine back to leucine represents the first sense-to-sense genetic code alteration engineered in any organism. Infiltration of natural and synthetic amino acids at nonsense codons and the engineering of *E. coli* strains and mammalian cells that tolerate high levels of amino acids misincorporation have been achieved and proteins with novel functionalities have been produced (Nangle *et al.*, 2006;Young and Schultz, 2010;Hoesl and Budisa, 2012). However, the main focus of these studies were the production and characterization of novel proteins containing fluorescent and other amino acid chemistries rather than cellular and evolutionary analyses (Wang *et al.*, 2009).

It is generally assumed that a sense codon reassignment depends in large part on the tolerance of the proteome to the reassignment. Accordingly, indiscriminate misincorporation of leucine at CUG positions in *C. albicans* affects more than half of its proteome (Gomes *et al.*, 2007), generating an array of protein variants in each cell (statistical proteome). Hence, the *C. albicans* proteome is plastic and highly complex. Furthermore, protein crystallography and molecular modeling showed that *C. albicans* tolerates global misincorporation at CUG codons due to the exquisite positioning of these codons at sites where incorporation of both leucine and serine have similar impacts on protein structure (Rocha *et al.*, 2011). Also, the atypical serine misincorporation evolved gradually over 100 million years and erased almost all (approximately 30.000) CUG codons from the genome of the CTG clade ancestor, creating a singular evolutionary condition where CUGs re-emerged through mutation of one of the six serine codons and codons coding for amino acids with similar chemical properties to serine (Massey *et al.*, 2003;Santos *et al.*, 2004). Altogether, these studies help explaining the high tolerance of *C. albicans* to increased ambiguity levels and especially the complete reversion of CUG identity in this organism.

In the present study we further show that CUG ambiguity is an important phenotypic diversity generator which is in line with previous data from our laboratory (Santos *et al.*, 1999;Miranda *et al.*, 2007). Besides evidences from yeast and fungi, there are numerous documented cases where a certain level of errors during translation can be tolerated and even generate advantageous phenotypes (Pezo *et al.*, 2004;Bacher *et al.*, 2007;Ruan *et al.*, 2008). For example, *Mycoplasma* parasites have error-prone AARSs: LeuRS, PheRS and PheRS with mutations and deletions in their editing domains (Li *et al.*, 2011). The mischarging of tRNAs by these AARSs generates mistranslation, leading to the production of statistical proteins with Phe/Tyr, Leu/Met and Leu/Val substitutions. The high frequency of mistranslation may provide a mechanism for their remarkable phenotypic plasticity and allow these parasites to escape host defences (Li *et al.*, 2011). Another interesting case, that highlights positive aspects of mistranslation, concerns the mammalian MetRS that is modified under environmental stress. These modifications allow the misacylation of noncognate tRNAs with methionine

and consequently provide a response mechanism to protect cells against oxidative stress as methionine is a reactive oxygen species (ROS) scavenger (Netzer *et al.*, 2009). Our discovery that codon ambiguity generates phenotypic diversity of high adaptation potential in *C. albicans* shows unequivocally that the potentially lethal proteome chaos generated by codon ambiguity is not an impediment for selection of codon reassignment.

Carl Woese postulated in his Translation Error hypothesis for the evolution of the genetic code that ambiguous codons were common in the primordial translation system, that the proteins produced were “statistical proteins” and that ambiguous codons were easily reassignable (Woese, 1965;Woese *et al.*, 2000). Yarus and Schultz (Schultz and Yarus, 1994) proposed that such codon ambiguity could also explain the evolution of natural codon reassignments in extant organisms with highly complex genomes (Schultz and Yarus, 1996). Our data validates experimentally both hypotheses. In fact, some of the advantages produced by CUG ambiguity (Figure 2-10 and Figure 2-11) support the hypothesis that codon mistranslation enhances effects of environmental stress on the evolution of phenotypic diversity. Indeed, ambiguous cells are pre-adapted to stress and survive better than wild type cells under variable stress conditions, as is clearly the case in presence of antifungals (Figure 2-10B and Figure 2-11) and SDS (Figure 2-10C and Figure 2-11). The molecular mechanisms of such stress pre-adaptation are unclear, but mistranslation may trigger the general stress response and stress cross protection (Santos *et al.*, 1997;Santos *et al.*, 1999;Silva *et al.*, 2007). If so, one should be able to identify environmental conditions where mistranslation is favorable and use them in forced evolution experiments to gradually increase codon ambiguity and ultimately achieve codon reassignment. This would provide a powerful strategy for creating organisms with alternative genetic codes that could produce new proteins containing novel amino acid chemistries.

2.6 Conclusions

This study describes the first artificial alteration of the genetic code and provides possible ways of overcoming the negative impact of codon ambiguity, namely through generation of phenotypic diversity.

The natural codon ambiguity of *C. albicans* has so far been viewed as a nuisance of gene translation. However, CUG ambiguity and general mistranslation are modulated by several stressors that likely affect tRNA expression, modification, charging and ribosome translation. Our discovery that CUG ambiguity has direct impact on phenotypic diversity adds therefore a new dimension to the study of the molecular mechanisms of phenotypic variation and ecological adaptation. It will be most fascinating to unravel how CUG ambiguity generates phenotypic diversity. Moreover, it is clear from this study that the toxicity of codon ambiguity is not an obstacle to codon reassignment, supporting the premise that ambiguity plays a critical role in the evolution of genetic code alterations.

3. Transcriptional profile of *Candida albicans* strains with an altered genetic code

3.1 Abstract

Work from the previous chapter demonstrated that the human pathogen *Candida albicans* supports high levels of leucine misincorporation and it is even able to cope with the complete reassignment of CUG codons from serine to leucine. Mistranslation at CUG positions produces sub-populations of stress resistant cells and phenotypic variability. In order to elucidate the molecular nature of these phenotypes, we compared the transcriptomes of strains that incorporate different levels of leucine at CUG positions. Gene expression deregulation was observed in all strains, but the classical hallmarks of the general stress response observed in *Saccharomyces cerevisiae* are absent. Only the *C. albicans* strains that mistranslate at very high level mounted a specialized core stress response characterized by a subset of genes that responded in a stereotypical manner to the proteome instability. Overall, this study provides new data on how *C. albicans* responds to high-level mistranslation and how the CTG clade ancestor may have overcome the early stages of CUG identity change.

3.2 Introduction

The human fungal pathogen *Candida albicans* can colonize several niches within its human host and, in specific conditions, can cause oral and vaginal thrush, as well as more serious mucosal and systemic infections (Odds, 1994). Several properties of *C. albicans* contribute to the success of this organism as a pathogen, namely the adhesiveness to host cells, the ability to switch from a yeast form of growth to a filamentous form, secretion of degradative enzymes and resistance to oxidative stresses and other environmental challenges. Depending on the site of infection and the events associated with the process, *C. albicans* must employ specialized stress responses to rapidly adapt to the changing environment (Hube, 2004). Therefore, survival and adaptation require a prompt reprogramming of the organism's transcriptome in which genes that encode stress

protective or repair functions are induced (Estruch, 2000;Enjalbert *et al.*, 2003;Gasch, 2007). This rapid adaptation has been widely studied at the transcriptional level by several groups. Indeed, DNA microarrays have become routine tools for the study of *C. albicans* and other yeasts biology. Among such studies are the response to oxidative stressors (Enjalbert *et al.*, 2006;Wang *et al.*, 2006b;Znaidi *et al.*, 2009), osmotic and heat-shock (Enjalbert *et al.*, 2003), the response to macrophage engulfment (Lorenz *et al.*, 2004), to pH changes (Bensen *et al.*, 2004) and the response to nitrosative stress (Hromatka *et al.*, 2005). These studies exposed clear differences between the stress response of this fungal pathogen and those of budding and fission yeasts (Gasch and Werner-Washburne, 2002;Gasch, 2007). For example, when submitted to environmental changes that stimulate core stress responses, *C. albicans* does not have a common transcriptional response as in the cases of *S. cerevisiae* and *S. pombe* (Enjalbert *et al.*, 2003;Gasch, 2007). In turn, *C. albicans* mounts a specialized core stress response mediated by Hog1 that is significantly smaller than in *S. cerevisiae* and *S. pombe* (Smith *et al.*, 2004;Enjalbert *et al.*, 2006). Furthermore, homologues of the *S. cerevisiae* transcription factors Msn2 and Msn4, that are crucial in regulating the core stress response in this budding yeast, do not have a corresponding role in *C. albicans* (Nicholls *et al.*, 2004). Globally, data from several transcript profiling experiments have illustrated that distinct responses are stimulated depending on the environmental challenge, genetic modification or some other manipulation.

Translation is not an error free process and among the variety of errors that can affect protein synthesis are missense, nonsense, frameshifting and ribosome drop-off errors (Farabaugh and Bjork, 1999). These errors lead to the production of aberrant, misfolded and/or unfolded proteins that ultimately result in a deleterious effect to the cell. Thus, in order to minimize the mRNA mistranslation effects, eukaryotic organisms employ numerous mechanisms: the universally conserved heat-shock response, the ubiquitin-proteasome pathway (UPS), molecular chaperones, autophagy and the ER associated protein degradation pathway (ERAD) (Lindquist and Craig, 1988;Pickart and Cohen, 2004;Schroder and Kaufman, 2005;Bukau et al., 2006). These processes are highly dynamic and

reactive since they increase or decrease the transcription and translation of their components in response to specific proteome quality needs. Hence, a number of genes (chaperones and proteasome subunits) are constitutively expressed at high level, while others are expressed at low-level under normal conditions and are strongly induced upon stress (McClellan *et al.*, 2005).

Surprisingly, translational inaccuracy can be in the order of 10^{-5} in the yeast *S. cerevisiae* (Stansfield *et al.*, 1998) or 10^{-2} in bacterial cells of *B. subtilis* (Meyerovich *et al.*, 2010) and in some cases those errors can originate advantageous phenotypes (Santos *et al.*, 1999;Pezo *et al.*, 2004;Silva *et al.*, 2007;Bacher *et al.*, 2007;Netzer *et al.*, 2009). Our previous work with low-level mistranslating *S. cerevisiae* strains showed that ambiguity at the CUG codon increases tolerance to oxidative stress (H_2O_2), osmotic stress (NaCl) and also enhances resistance to heavy metals (arsenite, cadmium) and drugs (cycloheximide) (Santos *et al.*, 1997;Santos *et al.*, 1999). In this case, mistranslation generated stress cross-protection through overexpression of molecular chaperones and other stress genes that protect cells against those different stressors (Silva *et al.*, 2007).

In the fungal pathogen *C. albicans*, low-level mistranslation also generated extensive phenotypic diversity responsible for numerous adaptive advantages under challenging conditions (Gomes *et al.*, 2007;Miranda *et al.*, 2007). Gene expression profiling of those strains revealed that leucine misincorporation at CUGs up-regulated the expression of genes involved in cell adhesion and hyphal growth and the secretion of proteases and phospholipases, features associated with virulence and infection (Miranda *et al.*, 2007). On the other hand, there was no activation of the stress response in *C. albicans* which was in sharp contrast with its activation in *S. cerevisiae*. This first attempt to correlate gene expression variations with phenotypes performed by Miranda (2007) prompted us to determine the effects of the complete CUG reassignment in gene expression.

The intriguing observation that cells could survive to complete CUG reassignment (previous chapter) and display growth advantages in a variety of stress conditions raised a series of questions concerning the molecular mechanisms that support these phenotypes. Transcriptome profiling would provide

the first insight into gene expression patterns of high-level mistranslating cells. Despite the fact that transcriptional profiles were not always linked with the phenotypes observed, we were able to correlate deregulation of *MDR1*, *CDR1* and *SOD* genes with drug and oxidative stress tolerance. As expected, a core stress response was absent in *C. albicans* with low levels of ambiguity, but cells with high levels of mistranslation (more than 67%) triggered a stress response similar to the one mediated by the Hog1 stress-activated protein kinase in osmotic and oxidative stress responses. Finally, transcriptome profiling of the engineered clones showed minimal impact on deregulation of functional categories in the least ambiguous clones T0KO1 and T2KO2 (0.64% and 98.5% of leucine misincorporation, respectively), suggesting that *C. albicans* copes better with 0.64% or 98% of leucine incorporation at CUGs (lower ambiguity) than with 20% to 80% of leucine incorporation (high ambiguity).

3.3 Methods

3.3.1 Strains and growth conditions

C. albicans SN148 (arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm434/ura3Δ::imm434iro1Δ::imm434/iro1Δ::imm434) (Noble and Johnson, 2005) was grown at 30°C in YPD (2% glucose; 1% yeast extract, and 1% peptone). *C. albicans* strains T0, T1 and T2 were grown in minimal medium lacking uridine and containing 0.67% yeast nitrogen base without amino acids, 2% glucose, 2% agar and 100 µg/ml of the required amino acids. Heterozygous $tRNA^{Ser}_{CAG}$ knock-out strains (T0KO1, T1KO1 and T2KO1) were grown in minimal medium lacking uridine and arginine, while homozygous deletion cells (T2KO2) were grown in medium lacking uridine, arginine and histidine. Cells were grown aerobically at 30°C unless indicated otherwise. All *C. albicans* strains used in this study are listed in Table 3-1 below.

Table 3-1. *C. albicans* strains used in this study.

Strain	Genotype	% Leu incorporation
T0	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm434/ura3Δ::imm434 iro1Δ::imm434/iro1Δ::imm434 RPS1/rps1Δ::pUA709 (URA3)	1.45
T0KO1	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm434/ura3Δ::imm434 iro1Δ::imm434/iro1Δ::imm434 RPS1/rps1Δ::pUA709 (URA3) tSCAG/tscagΔ::ARG4	0.64
T1	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm434/ura3Δ::imm434 iro1Δ::imm434/iro1Δ::imm434 RPS1/rps1Δ::pUA702 (URA3, Sc tLCAG)	20.61
T1KO1	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm434/ura3Δ::imm434 iro1Δ::imm434/iro1Δ::imm434 RPS1/rps1Δ::pUA702 (URA3, Sc tLCAG) tSCAG/tscagΔ::ARG4	50.04
T2	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm434/ura3Δ::imm434 iro1Δ::imm434/iro1Δ::imm434 RPS1/rps1Δ::pUA706 (URA3, Sc tLCAG, Sc tLCAG)	67.29
T2KO1	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm434/ura3Δ::imm434 iro1Δ::imm434/iro1Δ::imm434 RPS1/rps1Δ::pUA706 (URA3, Sc tLCAG, Sc tLCAG) tSCAG/tscagΔ::ARG4	80.84
T2KO2	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm434/ura3Δ::imm434 iro1Δ::imm434/iro1Δ::imm434 RPS1/rps1Δ::pUA706 (URA3, Sc tLCAG, Sc tLCAG) tscagΔ::ARG4/tscagΔ::HIS1	98.46

3.3.2 Total RNA extraction

Total RNA was extracted using an acidic hot-phenol protocol (Schmitt *et al.*, 1990). About 50 ml of exponentially growing cells (OD_{600nm} = 0.5) were harvested by brief centrifugation at 4000 rpm for 3 minutes at room temperature. After quick removal of the supernatant, tubes were immediately immersed on liquid nitrogen and frozen at – 80ºC. Frozen pellets were ressuspended in 500 μl of acid phenol:chloroform (Sigma, 5:1, pH 4.7) kept at 65ºC. The same volume of TES-buffer (10 mM Tris pH 7.5, 10 mM EDTA, 0.5 % SDS) was added and the tubes were vortexed for 20 seconds to ressuspend the cell pellet. After 1 hour incubation in a water bath at 65ºC, with 20 seconds vortexing every 10 minutes, the tube content was transferred to a 1.5 ml eppendorf tube and centrifuged for 20 minutes

at 14000 rpm at 4ºC. The water-phase was added to a new eppendorf tube, filled with 500 μl of acid phenol:chloroform (Sigma, 5:1, pH 4.7), vortexed for 20 seconds and centrifuged for 10 minutes at 14000 rpm at 4ºC. The water-phase from this step was added to a new eppendorf tube filled with 500 μl chloroform:isoamyl-alcohol (Fluka, 25:1), vortexed for 20 seconds and centrifuged for 10 minutes at 14000 rpm at 4ºC. Again, the water phase was transferred to a new eppendorf tube with 50 μl sodium acetate (3 M, pH 5.2). The tube was filled with ethanol (100 %, kept at – 20ºC) and incubated at – 20ºC for 1 hour. After RNA precipitation, tubes were centrifuged for 10 minutes at room temperature, 14000 rpm. The liquid was removed carefully to avoid touching the RNA pellet. The pellet was washed with 500 μl ethanol (80 %, kept at – 20ºC) and centrifuged for 5 minutes at room temperature, 14000 rpm. After removal of all traces of ethanol, the RNA pellet was air dried for 5 minutes and dissolved in sterile (mQ) water to a concentration of approximately 10 μg / μl (100 μl per 25 mid-log OD units, corresponding to about 1 mg total RNA). In order to remove possible DNA contamination, total RNA samples were treated with DNase I (Amersham Biosciences), according with the commercial enzyme protocol. Samples were frozen using liquid nitrogen and kept at – 80ºC.

3.3.3 Reverse Transcription and cDNA labelling

RNA labelling was carried out using the protocols described in the Agilent Quick Amp Labeling Kit (from Agilent One-Color Microarray-Based Gene Expression Analysis Kit). Samples were prepared by mixing 600 ng of total RNA, the appropriate amount of the third dilution of the Spike-Mix and T7 Promoter primer (Agilent). This reaction was incubated at 65ºC for 10 minutes to denature the primer and the template. A cDNA master mix (containing First Strand Buffer, Dithiothreitol, dNTP mix, MMLV-RT and RNaseOut) was added according to the instructions of the kit. Samples were then incubated at 40 ºC for 2 hours followed by incubation at 65 ºC for 5 minutes. A transcription master mix (containing

Transcription Buffer, Dithiothreitol, NTP mix, PEG, RNaseOut, Inorganic Pyrophosphatase, T7 RNA Polymerase and Cyanine 3-CTP) was added according to the instructions of the kit. Samples were then incubated at 40 ºC for 2 hours.

Before hybridization, the amplified cRNA samples were purified using Qiagen's RNeasy mini spin columns according to the corresponding instructions. The cyanine 3 dye concentration (pmol/µL), RNA absorbance ratio (260 nm/280 nm) and cRNA concentration (ng/µL) values were measured using a Nanodrop spectrophotometer. Determination of the yield and specific activity of each reaction was performed as follows:

- cRNA yield
  (Concentration of cRNA) * 30 µL (elution volume) / 1000 = µg of cRNA

- Specific activity
  (Concentration of Cy3) / (Concentration of cRNA) * 1000 = pmol Cy3 per µg cRNA

3.3.4 Hybridization and microarray washing

To prepare hybridization, 1.65 µg of Cy3-labeled cRNAs were mixed with the fragmentation mix (Blocking Agent and Fragmentation Buffer) and incubation lasted for 30 minutes at 60 ºC. Finally, GEx Hybridization Buffer HI-RPM was added and the preparation was assembled in the custom made Agilent arrays (GE1_105_Dec08). Slides were prepared using Agilent gasket slides according to the manufacturer instructions. Each hybridization was carried out for 17 hours at 65ºC, in an Agilent hybridization oven. Hybridization for each strain was done in triplicate.

Hybridized microarrays were disassembled at room temperature using GE Wash Buffer 1. The first wash took 1 minute with GE Wash Buffer 1 and the second wash (GE Wash Buffer 2) took place at an elevated temperature for 1

minute. Slides were rapidly dried by centrifugation for 6 minutes at 800 rpm, room temperature. After washing and drying, the microarrays were immediately scanned using the Agilent G2565AA microarray scanner (Agilent) and raw data was extracted using the Agilent Feature Extraction Software.

The microarrays used in this study were designed from assembly 21 of the *C. albicans* genome using eArray from Agilent Technologies. A total of 6,089 genes are represented by two sets of probes, both spotted in duplicate. Probes are randomly distributed. Four copies of each array were printed on a single slide (4 × 44,000) and hybridized individually.

3.3.5 DNA microarray data analysis

Analysis of raw data was performed using BRB-ArrayTools v3.4.0 developed by Dr. Richard Simon and BRB-ArrayTools Development Team (Simon *et al.*, 2007). BRB-Array Tools incorporates the Bioconductor R functions and the R programming language required for raw data normalization within arrays (Gentleman *et al.*, 2004). Each gene's measured intensity was median normalized to correct for differences in the labelling efficiency between samples. Median normalized values were then divided by its value in the control sample. Finally, the $\log_2$ values of the ratios (M values) for each triplicate were averaged. Statistical analysis, heatmaps and clustering of genes were performed using MeV software (Saeed *et al.*, 2006). Genes were identified as differentially expressed (DEGs) using the following criteria:

- Test-design: between subjects
- Variance assumption: Welch approximation
- P-value parameters: p-values based on t-distribution; alpha critical p-value =0.01
- False discovery corrections: just alpha

3.3.6 Gene expression deregulation analysis

Differentially expressed genes (DEGs) for each mistranslation strain were extracted using MeV software (Saeed *et al.*, 2006), considering as differentially expressed a variation equal or higher than 1.5X between each strain and the control T0. Only genes with significance level below an alpha corrected p-value of 10^{-3} were considered as differentially expressed. Hierarchical clustering of deregulated genes was carried out using MEV software using the Pearson correlation distance (Eisen *et al.*, 1998).

GO term enrichment for DEGs listed was carried out using the *Go Term Finder* tool in the *Candida Genome Database* (Skrzypek *et al.*, 2010), applied over GO biological process, and then selecting GO terms enriched with a p-value lower than 10^{-3}. The hierarchical clustering was carried out by constructing an expression matrix containing the category profiles of genes annotated within the enriched GO terms. The values of this matrix were averaged by GO term and mistranslating strains were clustered accordingly.

3.3.7 35S-Methionine incorporation assay

To quantify the rate of protein synthesis, a pulse-labeling experiment was used according to the protocol of Schieltz (1999). Radiolabeling of yeast proteins with L-[35S]-methionine and their analysis by polyacrylamide gel electrophoresis (PAGE) allowed for sensitive detection of both high and low abundance proteins. Exponentially growing cells (OD_{600nm} = 0.5) were calculated using a Newbauer chamber. $2x10^8$ cells were harvested by centrifugation at 4000 rpm for 5 minutes at room temperature. After a quick removal of the supernatant, pellet was resuspended in media without methionine (pre-warmed at 30ºC) and incubated for 20 minutes at 30ºC and 180 rpm. Hot methionine [35-S] (10.2 mCi/ml) was added and the mixture was incubated for 8 minutes at 30ºC. Protein synthesis was inactivated by adding the protein synthesis inhibitor cycloheximide, to a final

concentration of 50 pg/mL, together with crushed ice. Cells were harvested (4000 rpm at 4°C for 5 minutes), washed twice with 20 mM phosphate buffer, pH 7.0, and stored as cell pellets at -80°C (Schieltz, 1999).

Pellets were resuspended in 0.3 ml of lysis buffer (50 mM PBS pH 7.0, 1 mM EDTA, 5% Glycerol, 1mM PMSF and EDTA-free protease inhibitors from Roche). The suspension was kept on ice, transferred to cryo-tubes and 1 volume of glass beads was added. The cell walls were digested in a Pre-cellys disruptor using 5 cycles of 10 seconds in the beater and 2 minutes on ice. Tubes were then centrifuged during 5 minutes at 13000 rpm and supernatants were removed into new tubes. This suspension was span down again for 10 minutes at 13000rpm and a clear protein extract was obtained. Protein quantification was carried out using the Pierce BCA Protein Assay Kit. Aliquots were fractionated on a denaturing 15% SDS-polyacrylamide gel and the gel was exposed overnight with intensifying screens and developed using a Molecular Imager FX (Biorad). For radioactivity counting, 20 µl of labelled protein were applied on 1cm^2 microfiber filters (GF/C, Whatman) and allowed to dry at room temperature. Counting was performed in a scintillation counter.

3.3.8 Analysis of the distribution of CUG codons present in deregulated genes

Differentially expressed genes (DEGs) lists were analysed for their CUG content. Annotated ORFs from each list were downloaded from the *Candida* Genome Database (d'Enfert *et al.*, 2005) and analyzed with ANACONDA (Moura *et al.*, 2005). This in-house built software package counted all CUG codons present in each annotated ORF.

3.4 Results

Reversion of CUG identity deregulates gene expression. In the previous chapter, we constructed a series of *C. albicans* strains that mistranslate from 0.64% up to 98.46% (Table 3-1). The strains with high mistranslation levels produced an array of morphological phenotypes and some of them were tolerant to antifungals and other stress agents suggesting that they might up-regulate stress genes. In order to elucidate if the phenotypic variation observed was associated with the gene expression deregulation caused by the codon reassignment, the transcriptional response to different levels of mistranslation was investigated by mRNA profiling of newly constructed *C. albicans* strains (Table 3-1). For this, total RNA was extracted from 3 different clones of each strain and was reverse transcribed. The corresponding cDNAs were labelled with Cy3 and hybridized to "custom-made" *C. albicans* arrays (kindly provided by Geraldine Butler). These one-color arrays permitted the analysis of the expression level of 6131 genes encoded by the *C. albicans* genome, using a significant number of biological replicates (n=3-7). For the data analysis, we compared the transcriptional profiles of independent cultures of *C. albicans* cells grown in YPD at 30ºC with the profile of naturally ambiguous cells T0 (reference sample).

From the data extracted from 25 hybridization experiments, we selected genes that qualified as "significantly deregulated" by passing both a statistical (t-test, $p<0.01$) and a fold-variation (1.5 fold up or down) cutoff. As shown in Figure 3-1, there was only a minor deregulation of gene expression (approximately 2.9% of the total number of genes analyzed in the array) in T0KO1 cells. These cells have the lowest level of leucine misincorporation, a value that is very close to T0 control cells, which can explain the poor deregulation. The profile of T0KO1 revealed 97 ORFs that were up-regulated and 83 ORFs that were down-regulated. With increasing levels of ambiguity at the CUG codon, gene deregulation increased to 16.9%-26.2% of the total number of genes analyzed in the array. T1 strain had 26.2% deregulated genes, among which 818 were up-regulated and 790 were down-regulated. In T1KO1 cells, 1039 genes were deregulated: 558 genes showed up-regulation of >1.5-fold, while 481 genes were down-regulated.

Interestingly, strains T2 and T2KO1 showed higher percentage of down-regulated than up-regulated genes (12.6% and 13.1% against 12.2% and 12.5%, respectively). Finally, reverted cells T2KO2 deregulated 19.7% of the genes corresponding to 622 up-regulated and 584 down-regulated genes (Figure 3-1).

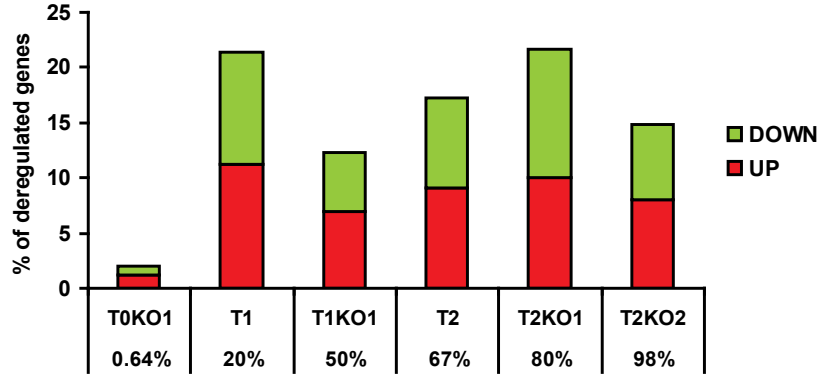


Figure 3-1. The mRNA profiling of the mistranslation set. Data showed significant alteration in gene expression in the majority of strains, the exception being T0KO1.

A detailed analysis of complete lists of deregulated genes (provided in annex F) showed that only 6 were commonly up-regulated in all mistranslating strains; 3 are uncharacterized and the remaining 3 are *ARG8*, *CPA1* and *ECM42*, which are involved in amino acid metabolism (arginine and glutamine). This suggested that manipulation of the genetic code has a direct impact on amino acid metabolism. On the other hand, *OYE23* was the only down-regulated gene in all strains. This gene encodes a putative NAPDH dehydrogenase which is involved in energy and general metabolism (Nett *et al.*, 2009).

As shown in Figure 3-2, results were organized by hierarchical clustering (Eisen *et al.*, 1998). On the y-axis, the deregulated genes were clustered according to similarity in their expression profiles (up-regulated genes are in red and down-regulated genes are in green). On the x-axis, mistranslating strains are organized on a growing level of leucine misincorporation. Genes in cluster 1 increased their expression until the 50% of ambiguity (T0KO1 < T1 < T1KO1) but their expression was down-regulated in strains with higher levels of mistranslation

(T2 < T2KO1 < T2KO2). The opposite situation was observed in cluster 4 where genes were down-regulated in strains with 50% of ambiguity and up-regulated at high levels of leucine misincorporation. Clusters 2 and 3 showed homogeneous patterns of expression across the entire mistranslating set: genes were down-regulated in the first cluster and up-regulated in the second cluster.

Overall, microarray profiling uncovered alterations in the expression of genes in strains mistranslating more than 20%. Strain T0KO1 did not show a significant deregulation of genes when compared to T0.

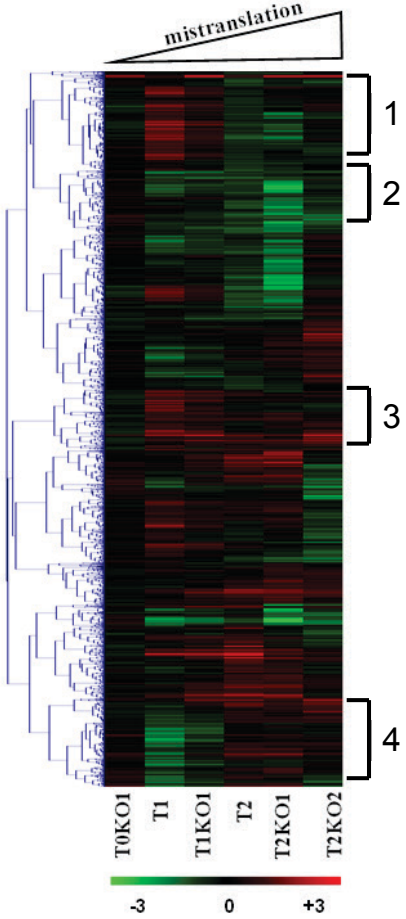


Figure 3-2. Schematic representation of the clustering of genes whose expression was deregulated by leucine misincorporation. A 0.6% to 98% CUG mistranslation set is represented. Gene expression values are normalized M-values against T0 (t-test between subjects using Welch approximation and p-value<0.01). Only genes with M-values corresponding to fold changes lower than -1.5 or higher than 1.5 in at least on strain of the spectrum are shown.

CUG codon reassignment deregulated global protein synthesis. All significant deregulated genes clustered in Figure 3-2 were submitted to a functional class analysis. In order to do the analysis, we used the *Go Term Finder* tool available at the Candida Genome Database with a p-value cutoff of 0.01. The distribution of up-regulated and down-regulated genes and their biological roles are shown in Figure 3-3 and Figure 3-4, respectively.

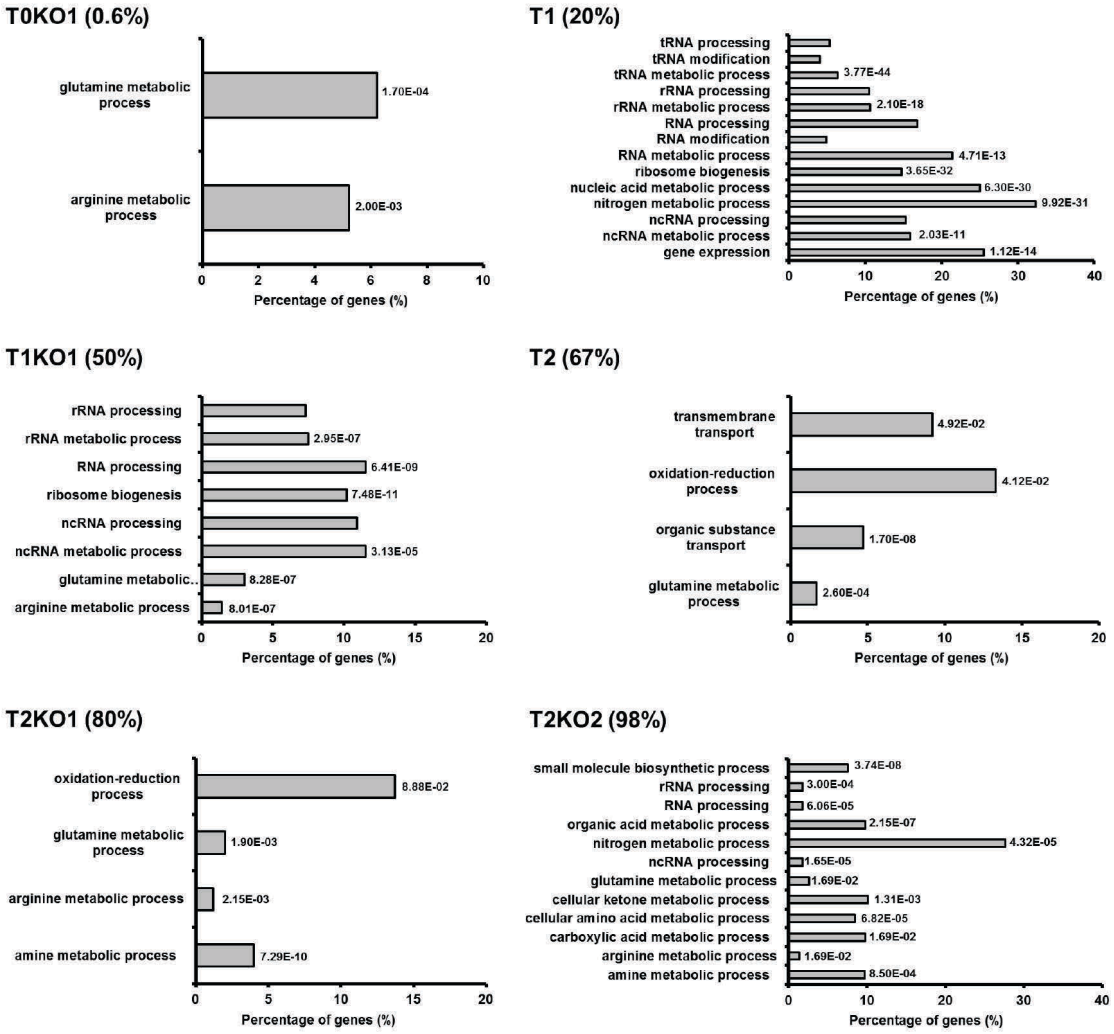


Figure 3-3. Functional class enrichment analysis of genes that are up-regulated by different levels of leucine misincorporation. Significantly up-regulated genes (fold-change > 1.5) were functionally enriched using the Go Term Finder tool (http://www.candidagenome.org/cgibin/GO/goTermFinder) available in the Candida Genome Database (p-value < 0.01). The p-value of enrichment for each GO category is indicated.

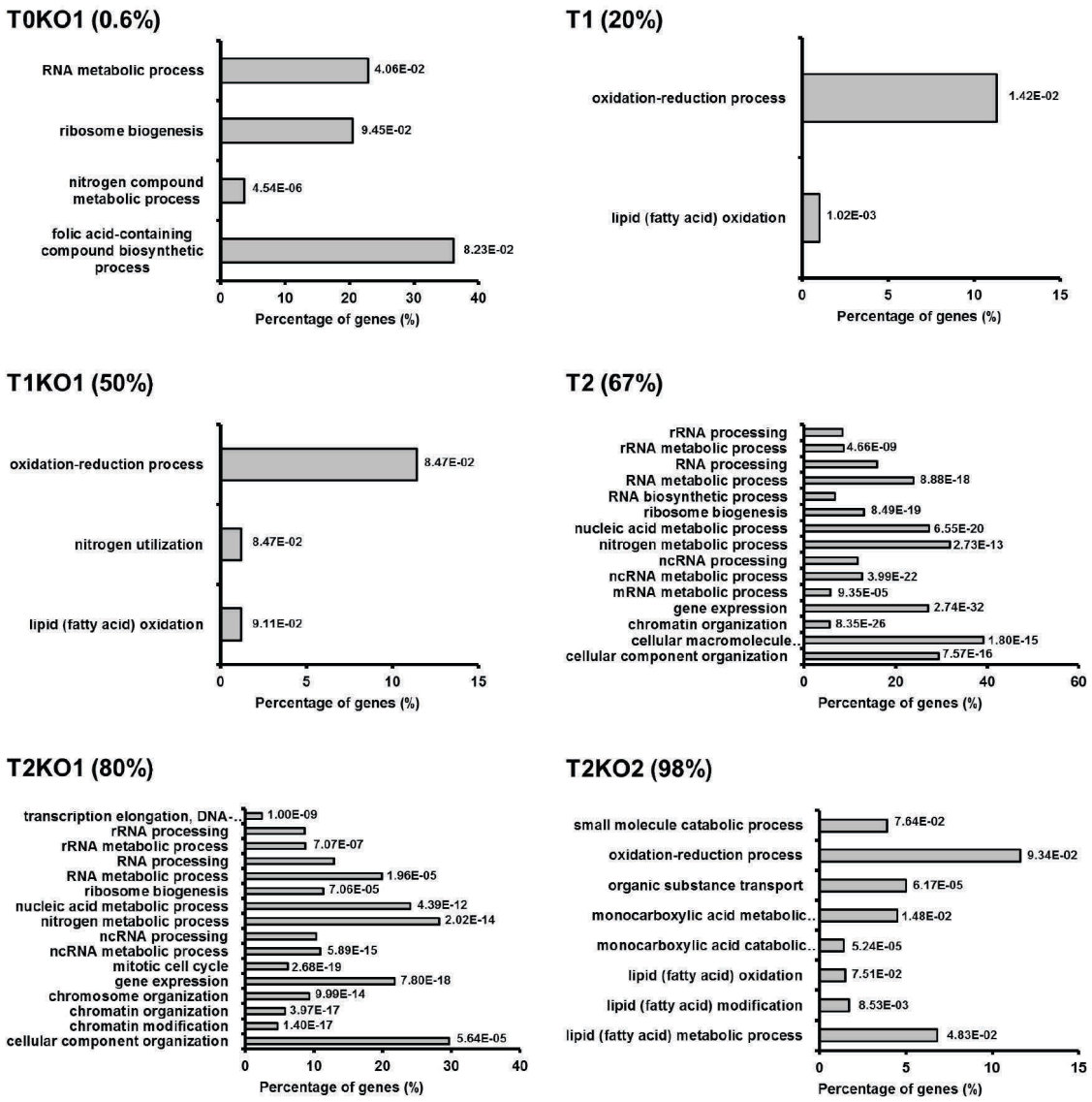


Figure 3-4. Functional class enrichment analysis of genes that are down-regulated by different levels of leucine misincorporation. Significantly down-regulated genes (fold-change < -1.5) were functionally enriched using the Go Term Finder tool (http://www.candidagenome.org/cgibin/GO/goTermFinder) available in the Candida Genome Database (p-value < 0.01). The p-value of enrichment for each GO category is indicated.

Surprisingly, transcriptome profiling of the engineered clones showed that only a limited number of functional categories were deregulated. In all cases the majority of responses involved genes of unknown function. As expected, the lowest functional enrichment of up-regulated genes was observed in strain T0KO1 where only 2 classes of genes were up-regulated, followed by T2 and T2KO1 with 4 classes. In these cases, misincorporation of leucine produced similar functional class enrichment concerning mostly amino acid metabolic processes (glutamine and arginine). Another category that was up-regulated in T2 and T2KO1 was oxidation-reduction (Figure 3-3 A, D, E). On the other hand, the number of up-regulated categories in strains T1, T1KO1 and T2KO2 was higher but this can be explained by the fact that most of those categories corresponded to groups of co-regulated genes involved in making the ribosome (Ihmels *et al.*, 2004;Hogues *et al.*, 2008;Lavoie *et al.*, 2009). One of these groups included genes that code for the ribosomal proteins (ribosome biogenesis) and a second group was composed of genes that assist in the proper assembly of the ribosome, such as genes involved in the processing of rRNA. The enrichment analysis showed that these categories had a high percentage of genes differentially expressed and p-values associated were exceptionally below 10^{-3} (Figure 3-3 B, C, F). Moreover, it should be noted the up-regulation of genes involved in tRNA processing and modification (T1) and genes of amino acid metabolic processes (T1KO1 and T2KO2).

Functional class analysis of the down-regulated gene dataset showed that strains T0KO1, T1 and T1KO1 had few functional categories affected. In other words, strains that incorporate less than 50% of leucine at CUGs showed little down-regulation of functional gene categories, while strains that misincorporate more than 50% of leucine (T2, T2KO1, T2KO2) showed the highest functional enrichment of down-regulated genes. Strains T1, T1KO1 and T2KO2 displayed common down-regulation of genes involved in oxidation-reduction processes and in the lipid and fatty acid metabolic processes (Figure 3-4 B, C, F). The latter is of particular importance since the process of morphogenesis and phenotype switching is dependent upon the lipid composition of *C. albicans* (Brown and London, 1998;Berman and Sudbery, 2002). Indeed, Goyal and Khuller (1994)

showed variations in the lipid composition of yeast and mycelia forms and proved that lipids constitute about 3.8-4.3 % of the dry weight of the fungal cell and are important structural and functional molecules in *C. albicans* (Goyal and Khuller, 1994).

As in the case of the up-regulated dataset, higher level of down-regulated categories corresponded to groups of co-regulated genes, as was the case in strains T0KO1, T2 and T2KO1 whose deregulated genes belong to gene expression, non-coding RNA metabolic process, ribosome biogenesis and rRNA processing. The enrichment analysis showed that these categories had a high percentage of differentially expressed genes and the p-values associated with them were lower than 10^{-3} (Figure 3-4 A, D, E). Additionally, genes involved in chromatin and chromosome organization (T2 and T2KO1) and genes of the mitotic cell cycle (T2KO1), which are important for the plasticity of the *C. albicans* genome and adaptation, were also down-regulated (Rustchenko, 2007;Ahmad *et al.*, 2008;Selmecki *et al.*, 2010).

It is important to notice the high deregulation of categories involved in growth-related processes, various aspects of RNA metabolism, namely RNA processing, translation elongation, tRNA processing and protein synthesis related processes. The intriguing aspect of gene deregulation in these categories is the trend of up-regulation in the strains that misincorporate leucine up to 50% (T1 < T1KO1), and the down-regulation trend in strains that misincorporate leucine at levels of 67% and 80% (T2 < T2KO1) (Figure 3-5). Strains at the extremes of the mistranslation scale (T0KO1 and T2KO2) showed almost no deregulation of those groups of genes. This surprising observation indicates that mistranslation deregulated protein synthesis differentially depending on the level of leucine misincorporation. It also implies that high levels of ambiguity (T2 and T2KO1) are more detrimental than the complete reversion of the genetic code (T2KO2). These results prompted us to validate the expression of several ribosomal protein genes by RT-qPCR, but these experiments confirmed the deregulation trend detected by the microarrays (data not shown).

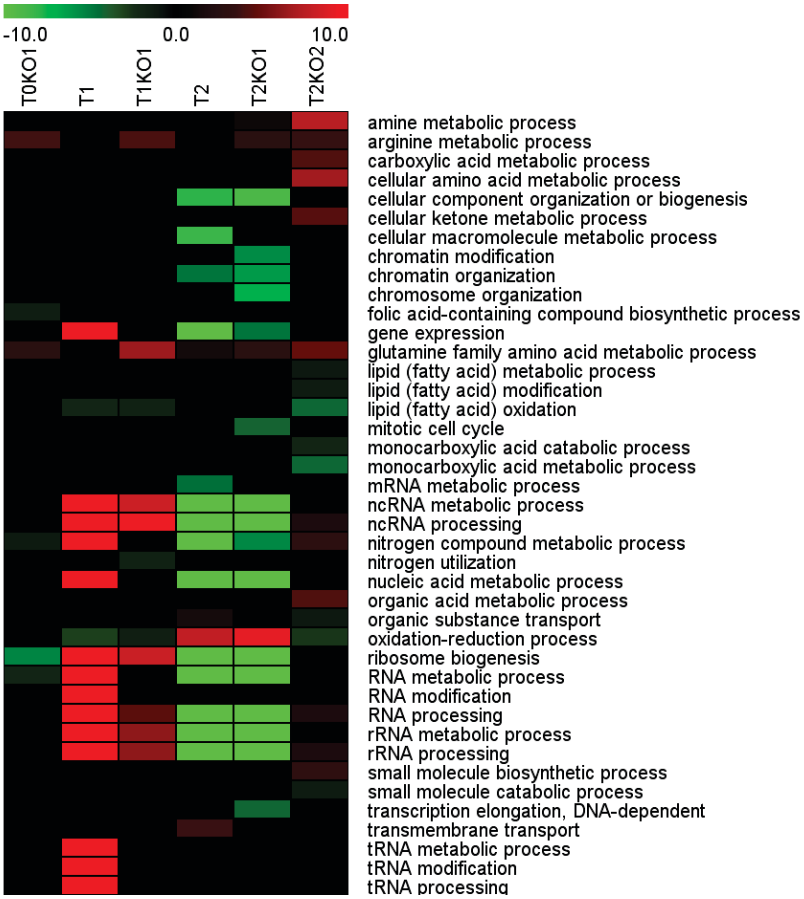


Figure 3-5. Summary of the GO terms extracted from differentially expressed genes lists. Each colour square represents the average expression level of the genes annotated with the corresponding GO term for each strain. Red means that genes within that category are up-regulated while green indicates down-regulation. Strains are hierarchically clustered in a growing level of leucine misincorporation. Significantly enriched functional classes were obtained using the Go Term Finder tool available in the Candida Genome Database with a corrected p-value cutoff of 0.01.

As mentioned above, strains incorporating less than 50% of leucine at CUGs showed clear up-regulation of the protein synthesis machinery, particularly genes involved in ribosome biogenesis and rRNA processing which suggests

increased protein synthesis rate in these strains (T1 and T1KO1). In order to determine whether this gene expression deregulation had a direct impact on protein synthesis rate, we pulse-labelled proteins by adding 1 µCi of [^{35}S]-methionine to 10^7 of actively growing cells. Protein synthesis was stopped by addition of cycloheximide and crushed ice. Surprisingly, strains T1 and T1KO1 showed a reduction of global protein synthesis rate relative to T0 (Figure 3-6). This paradoxical result highlights a gap between deregulation of the protein synthesis machinery and protein synthesis rate which should be further studied in the future. However, this did not happen in strains T2 and T2KO1 that incorporate more leucine than serine at CUGs (67% and 80%, respectively). In these strains the detected decrease of protein synthesis (Figure 3-6) was expected and correlated well with the down-regulation of genes involved in rRNA processing and ribosome biogenesis. Moreover, T2 and T2KO1 strains downregulated the expression of the *TBF1* gene by -2.38 and -2.48-fold, respectively. Tbf1 is essential, activates ribosomal protein gene expression and regulates the rDNA locus (Hogues *et al.*, 2008). Indeed, Tbf1 is the key DNA-binding element of the regulatory circuit controlling the expression of ribosomal proteins, a role analogous to that of the *S. cerevisiae* Rap1 (Lavoie *et al.*, 2009). Down-regulation of *TBF1* in our *C. albicans* strains can therefore be linked to the down-regulation of genes involved in ribosome biogenesis and rRNA processing and subsequently to the impaired global protein synthesis observed in T2 and T2KO1.

We were expecting that the strain T2KO2 would show the lowest rate of protein synthesis since it is decoding CUG codons as leucine only. However, this did not happen, rather the strain misincorporating 80% of leucine at CUGs was the one with lowest protein synthesis rate (Figure 3-6), which again suggests that *C. albicans* cells cope better with 100% than with 80% of leucine misincorporation.

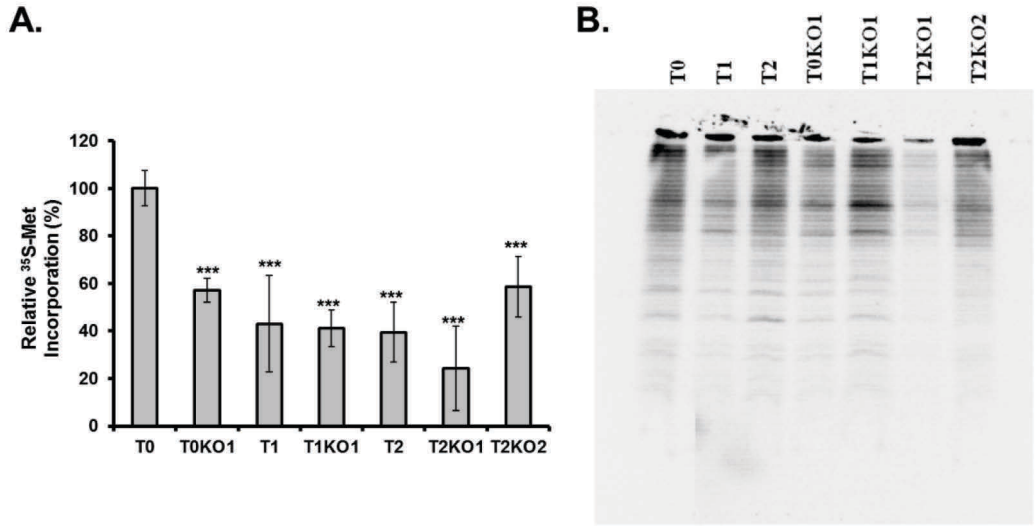


Figure 3-6. The effect of leucine misincorporation at CUGs on protein synthesis rate. A. Mistranslation decreased the rate of protein synthesis. Control T0 cells and mistranslating cells were pulse labelled with [^{35}S]-methionine for 20 minutes at 30ºC. Results are expressed as methionine incorporated per number of cells. One-way analysis of variance was applied with a Dunnett's Multiple Comparison Test against control T0 (***p-value<0.001). **B.** Labeled proteins (25µg) from each sample were fractioned using 12% SDS-PAGE and dried gels were exposed onto a phosphor screen for one week.

Leucine misincorporation at CUGs has little impact on the stress response. Previous studies carried out with *S. cerevisiae* showed that low-level mistranslation affects proteome stability and up-regulates molecular chaperones, protein degradation and unfolded protein responses (Silva *et al.*, 2007). Genes within these categories belong to the so-called environmental stress response (ESR) initially described in *S. cerevisiae*, which is characterized by the modulation of a general set of genes by numerous stresses, including heat-shock, oxidative stress, osmotic shock, nutrient starvation, DNA damage and extreme pH (Gasch *et al.*, 2000;Causton *et al.*, 2001;Gasch, 2007). This response includes the up-regulation of approximately 300 genes and down-regulation of nearly 600 genes. Up-regulated genes are related to carbohydrate metabolism, metabolite transport, fatty acid metabolism, detoxification of reactive oxygen species, autophagy, protein folding and degradation, cell wall modification, DNA-damage repair,

secretion, vacuolar and mithocondrial functions. Together with the induction of these stress-related genes, the ESR is characterized by the down-regulation of ribosomal proteins and ribosomal biogenesis genes (Gasch *et al.*, 2000;Causton *et al.*, 2001;Gasch and Werner-Washburne, 2002).

The analysis of the significantly enriched functional gene classes showed that the global response to mistranslation in *C. albicans* is very different from the yeast environmental stress response (ESR) or the response of *S. cerevisiae* to low level of mistranslation. In fact, in the list of *C. albicans* deregulated functional categories there was no category related to the typical stress response (Figure 3-3 and Figure 3-4), the exception being the fatty acid metabolism and ribosomal proteins genes. This data is in line with previous studies showing the inexistence of a general stress response in *C. albicans*. Indeed, heat shock (23ºC – 37ºC), osmotic shock (0.3 M NaCl) and oxidative stresses (0.4 mM H_2O_2) did not induce a general stress response in *C. albicans* (Enjalbert *et al.*, 2003). However, the global response of *C. albicans* to stressors that activate the Hog1 protein kinase pathway involve a small subset of genes that are commonly induced by 0.3M NaCl, 5 mM H_2O_2 and 0.5 mM $CdSO_4$ (Smith *et al.*, 2004;Enjalbert *et al.*, 2006). Nevertheless, the core stress response in this pathogen is significantly smaller than in other yeasts because only ~24 genes were up-regulated in the studies mentioned above. Most of those genes do not have a known function, while others are involved in oxidative stress (*CTA1*), cell wall biogenesis (*ECM41*), carbohydrate metabolism (*HXT61*, *HXT5*, *GLK1*, *GPD2* and *IPF20104*), multidrug resistance (*CDR4*) and protein folding and degradation (*RPN4* and *IPF17186*); others encode a copper-transport P-type ATPase (*CRD1*), an alkaline phosphatase (*IPF3094*), a mitochondrial respiratory function protein (*MRF1*), a reductase (*GRP2*) and a putative protein kinase predicted to regulate the activity of nitrogen source transporters (*NPR1*). These stress studies also unveiled down-regulation of 37 genes that are involved in protein synthesis and RNA processing (*IPF966*, *IPF3709*, *NOP4*, *NMD3*, *MRPL3* and *RCL1*), transport (*NMD5* and *PHO84*) and transcription (*RRN3*, *RPO41* and *IPF16752*) (Enjalbert *et al.*, 2006).

In our study, deregulated genes in strains at the end of the leucine misincorporation spectrum - T0KO1 (0.64%) and T2KO2 (98%) – and with ambiguity up to 50% – T1 (20%) and T1KO1 (50%) – showed no overlap with the list of modulated genes unveiled in the studies of Enjalbert and colleagues (2006). Remarkably, strains with more than 50% of ambiguity triggered a transcriptional response similar to the core stress response described by Enjalbert and colleagues (2006) and Smith and colleagues (2004), suggesting that T2 and T2KO1 cells activate the Hog1 pathway. Indeed, a comparison of the genes that were differentially expressed in T2 (67%) and T2KO1 (80%) and the core stress response described by Enjalbert and colleagues (2006) identified 16 and 17 genes with similar expression patterns, which corresponded to 26% and 28% of the *C. albicans* core stress response genes, respectively (Table 3-2). Cells incorporating 67% of leucine at CUG positions (T2 cells) up-regulated the GRP2 reductase and the mitochondrial respiratory function protein (*MRF1*), and down-regulated genes related to protein synthesis (*NOP4*) and transport (*PHO84*). The same categories were up-regulated in T2KO1 with the addition of the *RPN4* gene involved in protein folding and degradation.

These results further strengthen the hypothesis that *C. albicans* copes better with 0.64% or 98% of leucine incorporation at CUGs (lower ambiguity) than with 67% to 80% of leucine incorporation (higher ambiguity). This prompted us to investigate the induction of molecular chaperones in more detail. Surprisingly, there was no induction of Heat Shock Proteins (HSPs), the main category of molecular chaperones. The exception was *HSP31* that encodes a putative 30 kDa HSP with an unknown function, which was up-regulated in all strains, excluding strains at the very beginning and end of the mistranslation spectrum (T0KO1 and T2KO2, respectively).

Table 3-2. Core stress response genes (Hog1 pathway). Genes whose expression was altered by 67% (T2) and 80% (T2KO1) of ambiguity.

Gene (systematic name)	*C. albicans* core stress response (Hog1)	T2	T2KO1
DBP2	Down	-2.00	-2.92
DRS1	Down	-2.12	-1.52
IPF13756 (orf19.4479)	Down	-2.52	-3.49
IPF15217 (orf19.3778)	Down	-1.48	-
IPF16479 (orf19.2319)	Down	-2.39	-6.14
IPF16752 (orf19.2711)	Down	-1.24	-
IPF16944 (orf19.3287)	Down	-1.81	-2.18
IPF3709 (orf19.3015)	Down	-1.73	-1.44
IPF5533 (orf19.2183)	Down	-1.55	-1.73
IPF966 (orf19.7552)	Down	-	-3.96
MPP10	Down	-1.87	-
NOP4	Down	-2.41	-5.77
PHO84	Down	-	-3.93
PLB1	Down	-	-3.20
PWP2	Down	-1.73	-1.56
RCL1	Down	-1.56	-
CDR1	Up	1.97	1.96
GRP2	Up	1.70	1.53
IPF18207	Up	-	1.81
MRF1	Up	4.33	4.56
NRG1	Up	-	1.39
RPN4	Up	-	2.05
TPS2	Up	-	1.41

Unlike Heat Shock Proteins (HSPs), the list of genes encoding protein folding chaperones (beyond the HSPs class of molecular chaperones) (Table 3-3) showed an obvious trend of up-regulation in the reverted strain T2KO2. The CCT family of genes that encode putative cytosolic chaperonines was the clearest example (*CCT2*, *CCT3*, *CCT5* and *CCT7* were up-regulated only in T2KO2). Also, *HCH1* was up-regulated and the corresponding protein is similar to the *S. cerevisiae* Hch1, which is a regulator of heat shock protein Hsp90 (Yin *et al.*, 2004).

Table 3-3. Fold variation in gene expression of protein folding chaperones.

Gene	Systematic name	T0KO1	T1	T1KO1	T2	T2KO1	T2KO2
CCT2	orf19.1402	0	0	0	0	0	1.19
CCT3	orf19.4004	0	0	0	0	0	2.15
CCT5	orf19.2288	0	0	0	0	0	1.30
CCT7	orf19.3206	0	0	0	0	0	2.29
CNS1	orf19.6052	0	1.42	1.34	0	0	2.29
EGD1	orf19.1154	0	0	0	0	0	1.57
HCH1	orf19.3396	0	0	0	0	0	2.15
PGK1	orf19.3651	0	0	0	0	0	1.23

Table 3-4. Fold variation in expression of genes related to the proteasome activity.

Gene	Systematic name	T0KO1	T1	T1KO1	T2	T2KO1	T2KO2
BLM3	orf19.2182	0	0	0	0	1.45	0
DOA4	orf19.7207	0	0	0	0	1.42	0
PR26	orf19.5793	0	0	0	0	1.59	1.28
PRE1	orf19.4025	0	0	0	0	1.53	0
PRE5	orf19.7178	0	0	0	0	1.42	1.27
PRE8	orf19.7335	0	0	0	0	0	1.48
PUP3	orf19.1336	0	0	0	1.33	1.55	1.37
RPN10	orf19.4102	0	0	0	0	1.38	1.58
RPN4	orf19.1069	0	0	0	0	2.05	0
RPN5	orf19.4032	0	0	0	0	0	2.03
RPN8	orf19.3168	0	0	0	0	0	1.67
RPT6	orf19.3593	0	0	0	0	0	1.46
SCL1	orf19.5378	0	0	0	0	0	1.51

In the reverted strain (T2KO2), the same tendency for up-regulation was observed in genes related to proteasome activity (Table 3-4), namely *PRE*, *RPN* and *RPT* genes that encode proteasome subunits. Interestingly, strain T2KO1 with

80% of leucine misincorporation also induced these genes. Particularly, the putative transcription factor *RPN4* that stimulates expression of proteasome genes (Gasch *et al.*, 2004;Enjalbert *et al.*, 2006) was 2.05-fold inducted. These results suggested that protein degradation is an important consequence of the proteome destabilization induced by high-level mistranslation. Overall, the data showed that *C. albicans* is highly tolerant to proteome destabilization induced by leucine misincorporation at CUGs because expression of genes related to protein folding and degradation was only induced in strains that mistranslate above 80%.

Phenotypic effects of gene deregulation. In an attempt to associate deregulated genes with the phenotypes described in chapter 2, we have searched our transcriptional dataset for specific induction or repression of genes involved in responses to antifungals, oxidative, osmotic and heavy metal stresses (Enjalbert *et al.*, 2003;Enjalbert *et al.*, 2006;Wang *et al.*, 2006b;Kusch *et al.*, 2007;Yin *et al.*, 2009).

In our phenotypic assay, strains T2 and T2KO1 exhibited high level of tolerance to fluconazole (Figure 3-7A). This is of special interest because of the widespread use of this azole in the treatment of *C. albicans* infections. Numerous molecular mechanisms contributing to azole resistance have been identified so far, including the up-regulation of *ERG11*, a gene encoding the protein lanosterol demethylase. This enzyme belongs to the ergosterol biosynthetic pathway that produces the major sterol of the fungal cell membrane. Elevated expression of genes from this pathway results in an elevated ergosterol content in the plasma membrane which contributes to increased resistance against cell membrane disrupting chemicals such as fluconazole (Sanglard, 2002;Sanglard *et al.*, 2003b;MacCallum *et al.*, 2010). However, in our study none of the ergosterol pathway genes were deregulated, suggesting that the tolerance observed in our strains was not related to up-regulation of *ERG11*. Resistance has also been associated with increased export of azoles from the cell, frequently associated with the induction of efflux pumps, namely the ATP binding cassette transporters CDRs

(*CDR1* and *CDR2*) and the major facilitator *MDR1* (Lyons and White, 2000;Coste *et al.*, 2006;Watamoto *et al.*, 2011). As shown in Figure 3-7B, T2 cells showed 3.22 and 1.97-fold up-regulation of *MDR1* and *CDR1*, respectively. T2KO1 cells showed 3.96 and 1.96-fold up-regulation of the same efflux pumps which provide putative explanations to the tolerance to fluconazole. In these strains, the high-level of leucine mistranslation led to the constitutive expression of *MDR1* and *CDR1* which may allow their enhanced survival when exposed to fluconazole.

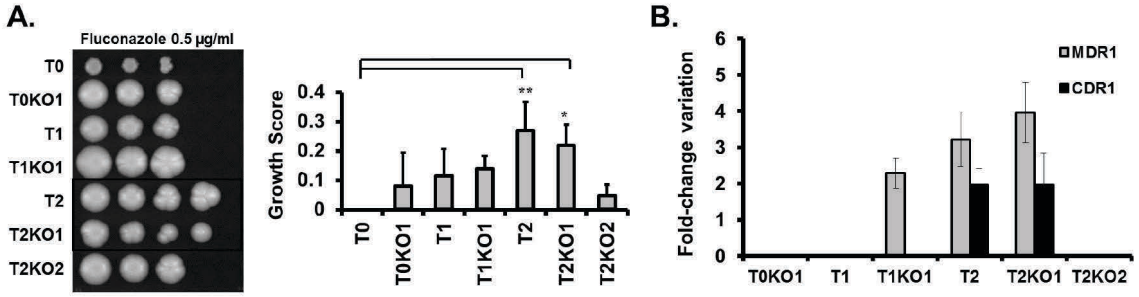


Figure 3-7. Tolerance of high mistranslating strains to fluconazole. A. Growth of *C. albicans* strains in YPD medium supplemented with 0.5 µg/ml of fluconazole. Strains T2 and T2KO1 showed a significant growth advantage when compared to the control T0. **B.** Fold variation in gene expression of the efflux pumps MDR1 and CDR1.

Another interesting result from the phenotypic assay described in the previous chapter was the statistically significant growth advantage of strain T2KO1 grown in media supplemented with menadione and hydrogen peroxide (H_2O_2) (Figure 3-8). Indeed, tolerance to oxidative stressors is of paramount importance since *C. albicans*, as any pathogenic fungi, has to cope with the oxidative burst generated by phagocytic cells of the host immune system. In order to explain the pre-adaptation in T2KO1 to oxidative stress, we have re-analysed our deregulated gene dataset and compared it to published transcriptional responses of *C. albicans* to exogenous oxidative stressors (Enjalbert *et al.*, 2003;Enjalbert *et al.*, 2006;Wang *et al.*, 2006b;Znaidi *et al.*, 2009) as well as proteomic analyses (Kusch *et al.*, 2007;Yin *et al.*, 2009). Remarkably, the T2KO1 strain displayed a

transcriptional response with similarities to the response of wild type *C. abicans* cells exposed to oxidative stressors in the mentioned studies. For example, T2KO1 cells activated the synthesis of detoxification mechanisms that included superoxide dismutases (*SOD1* and *SOD2*) and antioxidant proteins (*CIP1* and *OYE32*) (Table 3-5). Therefore, up-regulation of these genes may establish a connection between gene deregulation and the growth advantage phenotype of T2KO1 in menadione and H_2O_2. Despite the similarities between the response to high-level mistranslation and the response to oxidative stress, the former did not trigger up-regulation of catalase (*CTA1*), components of the thioredoxin (*TRR1* and *TRX1*) and glutaredoxin systems (*GPX1* and *GST3*) which constitute the most common response in *C. albicans*.

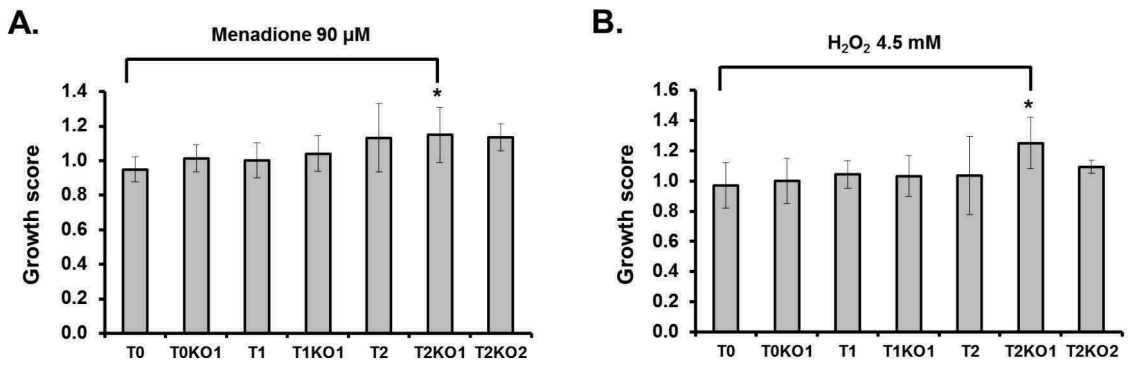


Figure 3-8. Tolerance of ambiguous *C. albicans* strains to menadione and hydrogen peroxide. A. Growth characteristics of *C. albicans* strains in YPD medium supplemented with 90 µM of menadione. Strain T2KO1 showed a significant growth advantage when compared to the control T0. **B.** Growth characteristics of *C. albicans* strains in YPD medium supplemented with 4.5 mM of H_2O_2. Strain T2KO1 showed a significant growth advantage when compared to T0. The growth score represents a ratio between growth in normal YPD medium and growth in YPD supplemented with the stressors. Data represent the mean ± standard deviation of triplicates of 3 independent clones (*p<0.1 one-way Anova post Dunnett's comparison test with CI of 95%, relative to the T0 control cells).

Table 3-5. Deregulation of oxidative stress genes.

Gene	Systematic name	T0KO1	T1	T1KO1	T2	T2KO1	T2KO2
CIP1	orf19.113	0.00	-2.61	3.42	2.43	5.47	2.04
OYE32	orf19.3131	0.00	-1.63	0.00	3.66	3.53	0.00
SOD1	orf19.2770.1	0.00	0.00	1.56	0.00	4.61	7.01
SOD2	orf19.3340	0.00	0.00	0.00	1.94	1.85	1.47
DDR48	orf19.4082	0.00	0.00	2.24	0.00	4.03	-5.45
ICL1	orf19.6844	5.19	2.06	2.24	1.50	1.97	0.00
PCK1	orf19.7514	0.00	-2.03	0.00	0.00	1.82	-3.03
CTA1	orf19.6229	0.00	0.00	0.00	0.00	0.00	-3.11
TRR1	orf19.4290	0.00	0.00	0.00	0.00	0.00	1.30
TRX1	orf19.7611	0.00	0.00	0.00	0.00	0.00	0.00
GPX1	orf19.87	0.00	0.00	1.39	0.00	0.00	2.11
GST3	orf19.8339	0.00	0.00	0.00	3.45	0.00	-1.48

Another interesting link between the phenotypic assay and gene deregulation could be established in the cases of the severe resistance of T2KO2 cells to $CuSO_4$ (Figure 3-9A) and sodium dodecyl sulfate (SDS). Both reagents are often utilized in phenotypic screens to test membrane integrity defects. SDS is known to cause membrane stress by solubilization of the phospholipids that make up most of the cell membrane. In the case of copper, there are various mechanisms of toxicity including plasma membrane remodeling and permeabilization resulting from membrane lipid peroxidation (Avery *et al.*, 1996). Remarkably, in T2KO2 cells there was an evident down-regulation of genes belonging to lipid metabolic process, modification and oxidation (Figure 3-5). Moreover, several studies showed that copper toxicity in *C. albicans* is prevented by ATPase transporters such as Crp1, Crd1 and Crd2, which remove metals from the cell, and a metallothionein named Cup1 (Riggle and Kumamoto, 2000;Weissman *et al*., 2000). In our data, transcript levels of genes *CUP1*, *CRD2* and *CRP1* showed increased expression in reverted cells (Figure 3-9B), particularly in the case of *CRD2* that had a 9.4-fold increase when compared to its expression in the control T0. Therefore, the constitutive upregulation of copper

transporters in T2KO2 cells may allow their enhanced survival when exposed to $CuSO_4$.

Despite the association of certain deregulated genes with various phenotypes, it was not possible to define the molecular mechanism of the pre-adaptations to stress that we observed. It is likely therefore that the phenotypic variability produced by genetic code ambiguity can only be explained by a combination of deregulated processes, namely hypermutagenesis or chromosomal rearrangements.

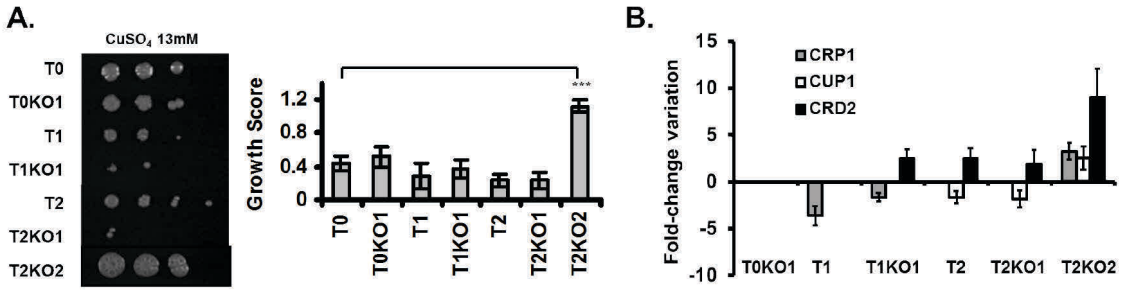


Figure 3-9. Tolerance of strain T2KO2 to copper. A. Growth characteristics of *C. albicans* strains in YPD medium supplemented with 13 mM of $CuSO_4$. Strain T2KO2 showed a significant growth advantage when compared to the control T0. **B.** Fold-change variation in expression of genes *CRP1*, *CUP1* and *CRD2* encoding proteins essential for copper resistance.

The effect of CUG usage bias on differential gene expression. A previous study showed that the *C. albicans* genome has 26148 CUGs distributed over 66% of its genes. The majority of those genes contain between 1 and 5 CUGs (57.7%) and the maximum number of CUGs within a gene is 38 (Gomes *et al.*, 2007). Since leucine misincorporation at CUGs could destabilize protein structure, we wondered whether our differentially expressed genes had a specific CUG usage bias. Thus, our dataset was analysed for their CUG content using the in-house built software ANACONDA (Moura *et al.*, 2005). The percentage of up-regulated genes with CUGs (Figure 3-10A) was similar to the global distribution of CUGs in

the *C. albicans* genome (66%). The exception was strain T0KO1 where less than 60% of up-regulated genes contained CUGs. In the down-regulated dataset, the majority of genes also had CUGs following the normal CUG distribution. The T1KO1 case was of special interest because 76% of its deregulated genes had CUG codons, which represented an enrichment of 10% relative to the normal distribution (Figure 3-10B). Despite these two exceptions, differential gene expression was not correlated with specific CUG enrichment.

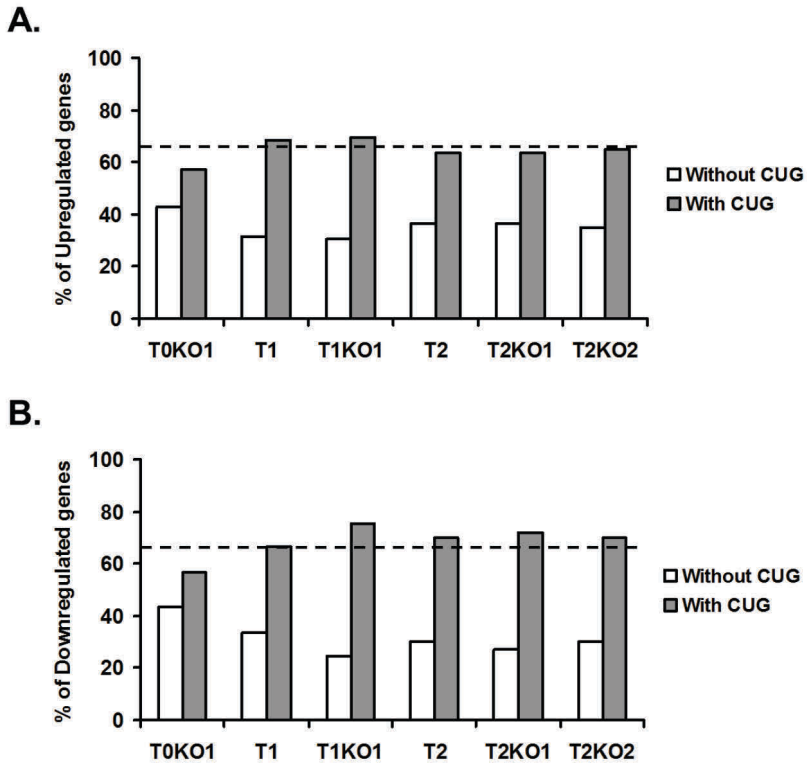


Figure 3-10. Distribution of CUG codons in *C. albicans* deregulated genes. A. Presence or absence of CUG codons in the up-regulated gene dataset of newly built mistranslating strains. **B.** Presence or absence of CUG codons in the down-regulated gene dataset of newly built mistranslating strains. The dashed line indicates the normal percentage of genes with CUGs within the *C. albicans* genome (66%).

Moreover, in naturally ambiguous strains, 83% of the highly expressed genes do not have CUGs which indicates that CUG usage is repressed in highly expressed genes. On the other hand, 81% of the genes expressed at low level have at least 1 CUG which reflects the relaxed CUG usage in genes whose expression is low (Santos *et al.*, 2011).

When we plotted the fold-change variation of each deregulated gene against its number of CUGs, we could not identify specific CUG usage biases in the deregulated genes. Up-regulated genes with higher fold variation had no CUGs and genes with CUGs had lower values of fold-change (Figure 3-11A). The same trend was observed in down-regulated genes (Figure 3-11B). Both distributions are indicated in Figure 3-11 for strain T1 but this trend was observed in all strains, indicating absence of CUG usage bias in the deregulated gene set.

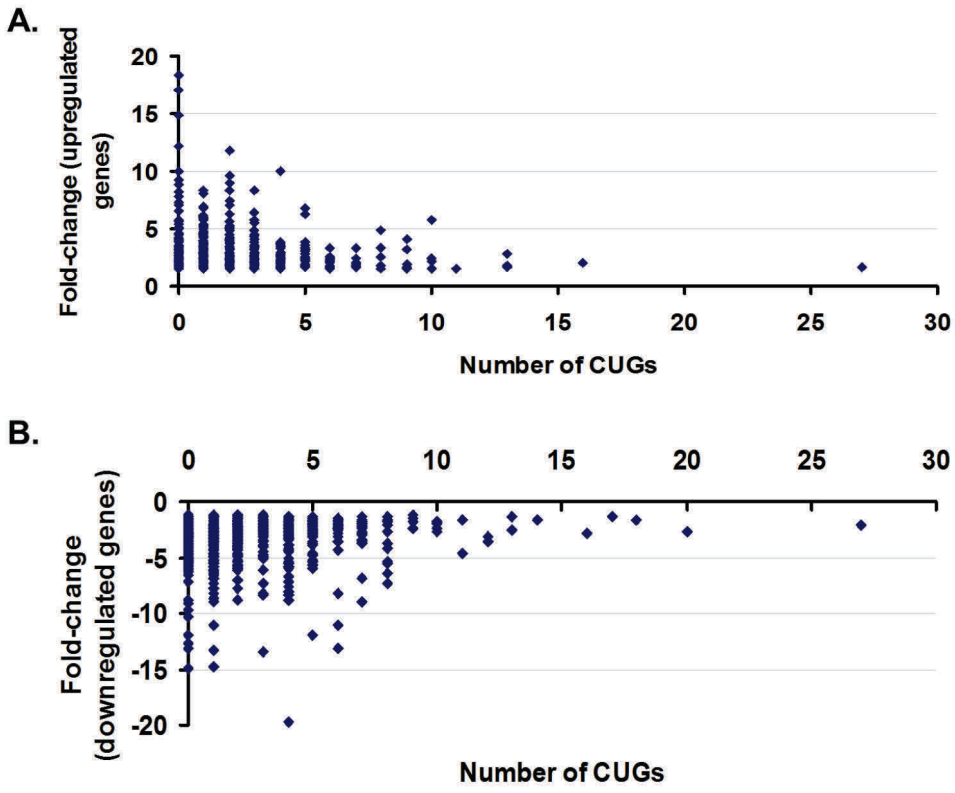


Figure 3-11. Analysis of the CUG content of deregulated genes. A. Distribution of fold-change variation in the up-regulated gene dataset of strain T1. **B.** Distribution of fold-change variation in the down-regulated gene dataset of strain T1.

3.5 Discussion

Due to the pathogenic nature of *C. albicans*, its ability to sense and promptly respond to changes in the environment is critical for survival. Therefore, several transcript profiling experiments have been performed in order to elucidate the responses stimulated by various environmental challenges (Lyons and White, 2000;Enjalbert *et al.*, 2003;Lorenz *et al.*, 2004;Bensen *et al.*, 2004;Hromatka *et al.*, 2005;Enjalbert *et al.*, 2006;Wang *et al.*, 2006b;Znaidi *et al.*, 2009). Although there is plenty of data concerning the molecular responses of *C. albicans* to a diverse range of environmental stresses, there is little information about the responses stimulated by genetic manipulations. This work attempted to fill that void by describing the transcriptional response of *C. albicans* mistranslating the genetic code. Strains with variable levels of leucine mistranslation (ranging from 0.64% to 80%) and a strain with reverted CUG identity were analyzed using DNA microarrays.

Taken together, our transcript profiling of *C. albicans* strains with increasing levels of leucine misincorporation identified 6 genes that were up-regulated and only 1 gene that was down-regulated across all strains. This highlighted the observation that the transcriptional response induced by low-level mistranslation is different from that induced by high-level mistranslation and that the metabolism of arginine and glutamine is affected in all strains.

Previous studies showed that yeasts exposed to severe stressors such as osmotic shock, changes of carbon sources, amino acids analogs and ethanol deregulate the expression of 10-20% of its genes (Trotter *et al.*, 2002;Berry and Gasch, 2008;Auesukaree *et al.*, 2009). In sharp contrast, minor stress deregulates only 1% of the yeast genes (Halbeisen and Gerber, 2009). In our work, high levels of mistranslation in strains T1, T1KO1, T2 and T2KO1 (20% to 80%) deregulated approximately 16.9%-26.2% of the total number of genes analysed, while low-level mistranslation (0.64% in strain T0KO1) deregulated only 2% of genes. This indicates that decreased leucine misincorporation produces a mild stress situation which is exacerbated in the cells that mistranslate above 20%. Also, transcriptome profiling of the engineered clones showed minimal impact on deregulation of

functional categories (Figure 3-5). The least ambiguous clones T0KO1 and T2KO2 (0.64% and 98.5% of leucine misincorporation, respectively) had a lower number of deregulated GO terms than the highly ambiguous clones (T1-20%, T1KO1-50%, T2-67% and T2KO1-80%), indicating that *C. albicans* copes better with 0.64% or 98% of leucine incorporation at CUGs (lower ambiguity) than with 20% to 80% of leucine incorporation (high ambiguity). The reason for this is unclear but one possibility if that protein-protein interactions involving mixtures of leucine (hydrophobic) and serine (polar) at CUG sites are less stable that those containing serine or leucine only. Supporting this hypothesis are results from Rocha and colleagues (2011) showing that CUG-encoded residues are predominantly located at the surface of *C. albicans* proteins (Rocha *et al.*, 2011). Hence, as most of the CUG-residues are exposed and consequently mediate macromolecular interactions, mixtures of leucine and serine at CUG sites should have greater impact in protein activity than serine or leucine only.

We have also identified a large cluster of genes that were coordinately up-regulated in the first part of the mistranslation spectrum (up to 50% of leucine misincorporation) and down-regulated in the strains that misincorporated leucine from 67% to 80%. The up-regulation pattern of these genes was then recaptured in the reverted strain. This cluster included ribosome biogenesis, processing and assembly, rRNA processing and RNA metabolic process which are related to growth and protein synthesis processes (Figure 3-5). The down-regulation of these categories in T2 (67%) and T2KO1 (80%) is consistent with the hypothesis that high levels of mistranslation have a strong negative impact on growth rate and on the rate of protein synthesis. This was evident in strain T2KO1 (80% mistranslation) which had the lowest protein synthesis rate (Figure 3-6) and 20% decrease in growth rate (previous chapter). In this context, it was intriguing that the same gene functional categories were up-regulated in strains T1 (20%), T1KO1 (50%) and in T2KO2 (98%). In particular, the latter strain showed the lowest growth rate (previous chapter) and it is generally assumed that ribosome biogenesis, protein synthesis and growth are coordinated. Indeed, rapidly growing yeast cells assign a large fraction of their biosynthetic capacity to ribosome

biogenesis (Levy *et al.*, 2007) and the number of ribosomes is often linked to cell size and growth rate (Jorgensen *et al*., 2004). Regardless of this generalized assumption, Levy and colleagues (2007) showed that ribosome biogenesis genes respond quickly to stress but the same genes are oblivious to long term changes in growth rate (Levy *et al.*, 2007). Indeed, those authors observed a lack of correlation between growth rate and expression of ribosome biogenesis genes in a compendium of 196 deletion yeast mutants, for which both growth rates and genome-wide expression profiles were reported (Hughes *et al.*, 2000a). Adaptation to mistranslation may therefore explain the pattern of ribosome gene deregulation observed in strains T1 (20%), T1KO1 (50%) and T2KO2 (98%).

Another result supporting the proposition that high-level codon ambiguity is more deleterious than the complete reversion of the genetic code was the down-regulation of *TBF1* and *CBF1* genes in strains T2 (67% mistranslation) and T2KO1 (80% mistranslation). These two genes belong to the transcriptional network controlling the ribosomal protein (RP) regulon that is essential to synchronize ribosomal biogenesis with cell growth and subsequently to coordinate ribosomal protein and rRNA levels allowing the continuous balance of ribosomal components. Hogues and colleagues (2008) used chromatin immunoprecipitation (ChIP) coupled with microarray analysis (ChIP-CHIP) in order to validate the interaction of Tbf1 and Cbf1 with RP gene promoters and they have discovered that Tbf1 is the main regulator of ribosomal protein genes while Cbf1 facilitates the access of Tbf1 to its targets. Therefore, down-regulation of *TBF1* and *CBF1* in T2 and T2KO1 strains might directly impact the regulation of genes involved in ribosome biogenesis and rRNA processing and subsequently the global protein synthesis of T2 and T2KO1. This is in line with the reduced growth and protein synthesis rates observed in these strains and with the observation that turning off *TBF1* expression caused a growth defect in the corresponding tbf1/TBF1 mutant strain (Hogues *et al.*, 2008). Also, microarray expression profiling of the tbf1/TBF1 mutant strain showed that ribosomal protein expression was decreased upon *TBF1* down-regulation which also happened in our T2 and T2KO1 strains.

The current model of reassignment of the CUG codon in Candida spp. postulates that the ancestor decoded the CUG codon as leucine using a single leucine tRNA ($tRNA_{CAG}^{Leu}$) and that the appearance of the mutant $tRNA_{CAG}^{Ser}$ was the critical step in the reassignment of CUGs (Massey *et al.*, 2003;Miranda *et al.*, 2006). This mutant serine tRNA could decode leucine CUG codons as serine (Santos and Tuite, 1995;Suzuki *et al.*, 1997) and competed for approximately 100 million years with the wild-type $tRNA_{CAG}^{Leu}$ for CUG decoding which resulted in a high ambiguity status of cells during that period. Finally, such ambiguity decreased over time due to the disappearance of the $tRNA_{CAG}^{Leu}$ gene (Massey *et al.*, 2003), however charging of the $tRNA_{CAG}^{Ser}$ with leucine and serine prevented complete change of identity of the CUG codon from leucine to serine (Santos *et al.*, 1996;Miranda *et al.*, 2006).

Our results showed that high levels of mistranslation are more detrimental to *C. albicans* cells than the complete reversion of CUG identity. This surprising result raised the question of how cells surpassed the early stages of CUG identity change where the level of ambiguity was high. Previous experiments from our group using genetically manipulated *S. cerevisiae* strains showed that codon-decoding ambiguity triggered the general stress response which created stress cross-protection and pre-adapted cells to environmental changes (Santos *et al.*, 1999;Silva *et al.*, 2007). In fact, numerous studies with *S. cerevisiae* and *S. pombe* showed that a large portion of their genomes respond to diverse stress conditions by changing the expression of a set of genes in a stereotypical manner (Gasch *et al.*, 2000;Causton *et al.*, 2001;Gasch and Werner-Washburne, 2002;Chen *et al.*, 2003;Gasch, 2007). This so-called environmental stress response (ESR) produces stress cross-protection, in which exposure to a specific type of stress can protect the cell against a subsequent exposure to an unrelated stress. In sharp contrast, *C. albicans* is not able to activate a global stress response (Enjalbert *et al.*, 2003;Nicholls *et al.*, 2004). Hence, the adaptive advantage of a general stress response could not explain the evolution of CUG reassignment in *C. albicans*. Recent work does suggest that *C. albicans* is able to activate a core stress response that also produces cross-protection (Smith *et al.*, 2004;Enjalbert *et al.*, 2006). The difference to previously described responses in *S. cerevisiae* and *S.*

pombe lies in the circuitry that controls the response. In other words, the core stress response in *C. albicans* evolved to fit the specific niche occupied by this human pathogen, i.e. it uses a different strategy to regulate its core stress response. In fact, the *Candida* lineage initiated more than 170 million years ago (Massey *et al.*, 2003) so there was extensive opportunity for divergence from other fungi such as *S. cerevisiae* and *S. pombe*. In the latter, the ESR is controlled by the Sty1 stress-activated protein kinase (SAPK) which is activated by oxidative and osmotic stresses, temperature changes, nutrient limitations and DNA damaging agents (Chen *et al.*, 2003). In *S. cerevisiae*, the ESR is not governed by a single regulatory pathway but by the transcription factors Msn2 and Msn4 through the general stress-responsive element (CCCCT) (Gasch *et al.*, 2000). In *C. albicans*, homologues of Sty1, Msn2 and Msn4 transcription factors have no role in the stress response (Nicholls *et al.*, 2004). That role is reserved for the Hog1 SAPK that is activated in response to osmotic, oxidative and heavy metal stressors (Smith *et al.*, 2004;Enjalbert *et al.*, 2006).

Only the strains that misincorporated leucine at 67% and 80% (T2 and T2KO1, respectively) exhibited typical transcriptional response activated by the Hog1 SAPK pathway. This result revealed that the core stress response was only activated at very high levels of ambiguity (more than 67%) which is in line with previous data showing that *C. albicans* tolerates well leucine misincorporation at CUGs. Indeed, deregulated genes from strains at the extremes of the leucine misincorporation spectrum - T0KO1 (0.64%) and T2KO2 (98%) – showed no overlap of genes modulated by environmental stressors (Enjalbert *et al.*, 2006). But, strains T2 (67%) and T2KO1 (80%) showed 26% and 28% of genes that overlap with the *C. albicans* core stress response genes, respectively (Table 3-2). Importantly, part of those genes belong to the core stress functional gene categories on *S. cerevisiae* and *S. pombe*, namely metabolite transport, protein synthesis, mitochondrial functions and protein folding and degradation. Altogether, our data suggested that mistranslation in *C. albicans* is somewhat toxic only in the cases of extreme CUG ambiguity.

Moreover, only the T2KO1 and T2KO2 strains up-regulated genes involved in protein folding and degradation. T2KO1 and T2KO2 up-regulated the *RPN4*

gene which encodes the transcription factor that regulates expression of proteasome subunit genes (Table 3-4) (Gasch *et al.*, 2004;Singh *et al.*, 2011). Also, T2KO2 up-regulated CCT genes that encode putative cytosolic chaperonines (Table 3-3). Up-regulation of these genes is usually associated with the degradation or recovery of aberrant proteins which is crucial for cells to survive proteome instability. Strains with less than 67% of ambiguity did not show this transcriptional response which is in line with previous results showing that increasing ambiguity (up to 28%) at CUG positions did not up-regulated proteasome activity and the heat shock protein response (Rocha *et al.*, 2011). This high tolerance for serine/leucine replacement was explained by the localization of 90% of the CUGs in non-conserved positions where accommodation of serine or leucine does not disrupt protein structure or function (Rocha *et al.*, 2011).

Taken together, the activation of a mild stress response in T2 (67%) and T2KO1 (80%) strains, and the up-regulation of protein folding and degradation genes in T2KO1 (80%) and T2KO2 (98%) indicates that CUG ambiguity is not neutral in *C. albicans*.

Evidence for positive outcomes of stress pre-adaptation in highly ambiguous *C. albicans* strains emerged by comparing gene deregulations with tolerance phenotypes (see chapter 2). For example, T2 cells up-regulated *MDR1* and *CDR1* genes that encode different types of efflux pumps that contribute to drug resistance (Lyons and White, 2000;Cannon *et al.*, 2007), explaining the growth advantage phenotype of T2 cells in media supplemented with fluconazole. Likewise, up-regulation of *SOD* genes (encoding ROS detoxifiers) in T2KO1 provided a clue for its tolerance to menadione and H_2O_2 (Herrero *et al.*, 2008;Frohner *et al.*, 2009), although this is not the most common response to oxidative stress in *C. albicans*. This may be linked to the fact that the Cap1 transcription factor has a key role in the oxidative stress regulation, independently of Hog1. The *CAP1* gene was not deregulated in T2KO1 and consequently genes from this regulatory pathway, namely *CTA1*, *TRR1*, *GPX1* and *GST3* that constitute the typical oxidative stress response in *C. albicans,* were not deregulated (Wang *et al.*, 2006b;Singh *et al.*, 2011).

Our gene expression data did not provide a clear explanation for the impressive phenotypic variation that we observed in the engineered strains, but supported the idea that mistranslating cells are pre-adapted to specific hostile environments. Additionally, the data suggested that chromosomal rearrangements and other genome alterations may also be linked to phenotypic variation. Indeed, the deregulation of genes involved in chromosome organization and mitotic cell cycle in T2KO1 cells (Figure 3-4E) support this hypothesis.

3.6 Conclusions

Our microarray analysis provided new information on how *C. albicans* deals with codon reassignment, in particular on how it modulates gene expression to mitigate the deleterious effects of such a dramatic event. Up until now, stress cross-protection did not appear to have a role in *Candida* evolution as this was not an apparent feature of *C. albicans* biology. The absence of a concerted stress response in *C. albicans* suggested that advantages associated to the general stress response could not explain the evolution of the nonstandard code in *C. albicans*. Our data show that *C. albicans* is highly tolerant to the manipulation of its genetic code and that it does not require activation of a stress response for CUG ambiguity levels below 50%. Cells incorporating more than 67% of leucine triggered a core stress response similar to that mediated by the Hog1 stress-activated protein kinase in osmotic and oxidative stress, indicating that CUG ambiguity toxicity only arises at very high level of leucine misincorporation, which may not be physiological.

Considering the close phylogenetic relationship between *S. cerevisiae* and *C. albicans*, our data support the hypothesis that the genetic code alteration may have remodelled the stress response of the latter, but this needs to be investigated in more detail in future studies.

4. Whole genome sequencing of engineered *Candida albicans* strains

4.1 Abstract

The genome flexibility of *Candida albicans* is believed to provide adaptation plasticity to sudden changes in environmental conditions, however it has been difficult to identify the exact genetic mechanisms underlying that adaptation plasticity. Next generation whole-genome sequencing allows for rapid determination of genome structure, including variation at the single nucleotide level, gene copy number alterations, loss of heterozygosity and aneuploidies. In this study, we used the Illumina Genome Analyser IIx platform to re-sequence more than 99.4% of the genome of five *C. albicans* strains with different levels of codon reassignment, with average alignment depths ranging from 56.8 to 96.2 per base. Both single nucleotide polymorphisms (SNPs) and loss of heterozygosity (LOH) events were detected. These genomic alterations accumulated preferentially in genes involved in filamentous growth, cell adhesion and drug resistance where CUG usage is positively biased, indicating that *C. albicans* uses its natural codon ambiguity to increase genomic diversity of relevance to pathogenesis and drug resistance processes.

4.2 Introduction

In recent years, a series of high-throughput DNA sequencing technologies have been developed which permit the annotation of complete genome sequences from many organisms, ranging from bacteria to humans (Kahvejian *et al.*, 2008;Shendure and Ji, 2008). Sequencing technologies such as 454/Roche, Illumina/Solexa, Helicos and ABI/SOLiD can produce gigabases of DNA sequence information in a single experiment which permits the assessment of genetic variation, RNA expression, protein-DNA interactions and chromosome structural changes associated with diseases, behaviour and other traits. These approaches have been applied to detect suppressor mutations in bacteria (Srivatsan *et al.*, 2008), to study the rate of evolutionary adaptation over 40000 generations in *E.*

coli (Barrick *et al.*, 2009), to reveal structural variation among all known human rhinovirus genomes (Palmenberg *et al.*, 2009), to characterize genetic variation in numerous human genomes (Pelak *et al.*, 2010) and to pinpoint alleles causing disease in exomes of affected individuals (Ng *et al.*, 2009;Ng *et al.*, 2010). Similarly, in yeasts there is an increasing number of studies concerning ecological and geographic distributions (Sampaio and Goncalves, 2008), genetic variation between wild and domestic populations (Liti *et al.*, 2009), sexual and/or asexual reproduction (Tsai *et al.*, 2008) and detection of adaptive mutations (Lynch *et al.*, 2008;Smith *et al.*, 2008;Otero *et al.*, 2010). Recently, the improvement of the genome sequences of three species of *Saccharomyces sensu stricto* yeasts (*S. bayanus* var. *uvarum*, *S. kudriavzevii* and *S. mikatae*) and their comparison to the genomes of *S. cerevisiae* and *S. paradoxus* enabled the determination of the branch-specific evolutionary rates for a set of 5261 protein-coding orthologs across the five *Saccharomyces* species and improved inferences of species-specific gains and losses (Scannell *et al.*, 2011). These evolutionary genetics studies proved that next-generation genomics technologies are crucial to understand genome evolution and the process of adaptation.

Various *Candida* genomes have also been sequenced and there are many resources available for comparative genomics of *Candida* species. Butler and colleagues (2009) reported the genome sequences of six *Candida* species and compared them to related pathogens and non-pathogens. The comparison exposed the evolutionary context of unusual features of *C. albicans*, namely cell wall gene family amplifications and the ability to mate using a parasexual cycle that lacks meiosis (Butler *et al.*, 2009). Some of these unusual characteristics are being intensively studied because they contribute to genome plasticity (Bouchonville *et al.*, 2009;Arbour *et al.*, 2009;Selmecki *et al.*, 2010;Forche *et al.*, 2011).

The *C. albicans* genome sequence was first published by the Stanford Genome Technology Center in 2004 (Jones *et al.*, 2004), and was annotated in 2005 (Braun *et al.*, 2005), but only in 2006 physical and genetic maps were used to produce a final chromosome assembly (Nantel, 2006). This physical map showed that the genome has 14.3-14.4 Mb (haploid genome) organized in 8

diploid chromosomes. *C. albicans* reproduces primarily through asexual clonal division (Graser *et al.*, 1996), although it has a mating locus in chromosome 5 (Hull and Johnson, 1999). Mating between diploid strains of opposite mating type can occur, which leads to the formation of tetraploids that do not undergo the regular meiotic reduction in ploidy (Magee and Magee, 2000;Bennett and Johnson, 2005). Instead, tetraploids undergo a "concerted chromosome loss", involving random loss of multiple chromosomes to a near-diploid state. The near-diploid progeny are often trisomic or homozygous for one or more chromosomes. Furthermore, numerous short range recombination events can occur within the parasexual progeny, yielding recombinant chromosomes (Bennett and Johnson, 2003;Forche *et al.*, 2008). These mitotic events are crucial to generate genetic diversity because meiosis has not yet been detected in *C. albicans*. The role of the parasexual cycle in the response to stress remains to be established but, as pinpointed in the work of Selmecki and colleagues (2010), the current model is that adaptive mutations evolve through somatic events, including point mutations, recombination, gene conversion, loss of heterozygosity, and/or aneuploidy (Bennett and Johnson, 2005;Selmecki *et al.*, 2010;Berman and Hadany, 2012).

Several studies established a correlation between genomic changes, including gross chromosomal rearrangements, aneuploidy, and loss of heterozygosity, and adaptive phenotypes in response to several perturbations (Rustchenko, 2007;Ahmad *et al.*, 2008). Rustchenko and co-workers found that monosomy of chromosome 5 allows for growth on sorbose and that the following reduplication of chromosome 5 under non-selective conditions reversed this phenotype, which results from the presence of multiple genes on the right arm of chromosome 5 that negatively regulate the *SOU1* gene (sorbose utilization 1) located in chromosome 4 (Kabir *et al.*, 2005). Aneuploidy is also prevalent in *C. albicans* strains that acquire resistance to fluconazole. A survey of resistant and sensitive strains found that ~50% of resistant strains carried at least one aneuploid chromosome and that ~20% of these strains had two extra copies of the left arm of the chromosome 5 organized as an isochromosome [i(5L)]. In these strains, resistance to fluconazole is a consequence of the increased number of copies of two genes on chromosome 5, namely *ERG11* (the drug target) and *TAC1*, a

transcription factor that up-regulates expression of the *CDR1* and *CDR2* efflux pumps (Selmecki *et al.*, 2006;Coste *et al.*, 2006;Selmecki *et al.*, 2009). Recently, Forche and colleagues (2011) measured the rates and types of loss of heterozygosity (LOH) events that appear in the absence and in the presence of several stressors, namely temperature, fluconazole and H_2O_2. This work showed that these stressors cause a significant increase in the rates of LOH and that this increase is proportional to the stress strength. Also, the types of LOH events that were detected differed in a stress-dependent manner, which indicates that eukaryotic cells generate dissimilar and increased genetic diversity in response to a variety of stress conditions (Forche *et al.*, 2011).

Here, we used the Illumina Genome Analyser IIx to re-sequence the whole genome of highly ambiguous and reverted *C. albicans* strains produced as described in chapter 2. We were able to generate high-depth sequencing data for the ambiguous T1, T2, T2KO1 and T2KO2 strains (with 20%, 67%, 80% and 98% of leucine misincorporation, respectively) and for the control genome T0 (wild type strain). Comparison of these genomes allowed us to uncover loss of heterozygosity events and single nucleotide polymorphisms (SNPs) in these mistranslating strains. An increasing number of genotype changes relative to the control strain at single nucleotide polymorphic sites with increasing degrees of leucine misincorporation were also observed. The strain that misincorporated 98.5% of leucine accumulated a high number of genotype changes and two strains that misincorporated more than 50% of leucine exhibited complete loss of heterozygosity in a large region of chromosome 5 covering genes related to drug resistance, filamentous growth and pathogenesis. These data highlight unanticipated roles of codon ambiguity in genome evolution and suggest that mistranslation generates genetic diversity, which in the case of *C. albicans* is associated with virulence and drug resistance.

4.3 Methods

4.3.1 Strains and growth conditions

Candida albicans SN148 (arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm434/ura3Δ::imm434iro1Δ::imm434/iro1Δ::imm434) (Noble and Johnson, 2005) was grown at 30°C in YPD (2% glucose; 1% yeast extract, and 1% peptone). *Candida albicans* strains T0, T1 and T2 were grown in minimal medium lacking uridine (0.67% yeast nitrogen base without amino acids, 2% glucose, 2% agar and 100 µg/ml of the required amino acids). Heterozygous $tRNA^{Ser}_{CAG}$ knock-out strains (T0KO1, T1KO1 and T2KO1) were grown in minimal medium lacking uridine and arginine, while homozygous deletion cells (T2KO2) were grown in medium lacking uridine, arginine and histidine. Cells were grown aerobically at 30°C unless indicated otherwise. All *C. albicans* strains used in this study are listed in Table 4-1 below.

Table 4-1. *C. albicans* strains used in this study.

Strain	Genotype	% Leu incorporation
T0	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm434/ura3Δ::imm434 iro1Δ::imm434/iro1Δ::imm434 RPS1/rps1Δ::pUA709 (URA3)	1.45
T1	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm434/ura3Δ::imm434 iro1Δ::imm434/iro1Δ::imm434 RPS1/rps1Δ::pUA702 (URA3, Sc tLCAG)	20.61
T2	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm434/ura3Δ::imm434 iro1Δ::imm434/iro1Δ::imm434 RPS1/rps1Δ::pUA706 (URA3, Sc tLCAG, Sc tLCAG)	67.29
T2KO1	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm434/ura3Δ::imm434 iro1Δ::imm434/iro1Δ::imm434 RPS1/rps1Δ::pUA706 (URA3, Sc tLCAG, Sc tLCAG) tSCAG/tscagΔ::ARG4	80.84
T2KO2	arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3Δ::imm434/ura3Δ::imm434 iro1Δ::imm434/iro1Δ::imm434 RPS1/rps1Δ::pUA706 (URA3, Sc tLCAG, Sc tLCAG) tscagΔ::ARG4/tscagΔ::HIS1	98.46

4.3.2 DNA content analysis

Strain ploidy was analysed by flow cytometry as described by Haase and Lew (1997). Briefly, cells from an overnight culture in YPD, at 30°C, were harvested by centrifugation at 10,000 *g*, washed twice in ice-cold MilliQ H_2O, counted using a hemacytometer and their concentrations adjusted to $1x10^7$ cells/ml. After centrifugation, cells were fixed overnight, at 4ºC, in 1 ml 70% ethanol (v/v), washed twice with wash solution A (50 mM Tris/HCl pH 8.0, 5 mM EDTA) and sonicated during the second wash for 5 seconds, followed by treatment with 2 mg/ml RNase (Sigma) in wash solution A for 16 h at 37°C. RNase-treated cells were resuspended in a 5 mg/ml pepsin solution in 55 mM HCl for 1 h at 37°C. Protease-treated cells were washed twice in wash solution B (50 mM Tris/HCl pH 7.5, 5 mM EDTA) followed by staining with 500 µl of 1 µM SYBR green I (Invitrogen) in wash solution B for at least 1 h at 4°C and in the dark. Cells were analyzed on a FACSCalibur (BD Biosciences, San Jose, CA) using CellQuest Pro software (BD Biosciences) to capture data (Haase and Lew, 1997).

4.3.3 Nuclear DNA staining with DAPI

To visualize nuclei, cells were grown to exponential phase in YPD and fixed in 70 % ethanol for 10 min. Fixed cells were washed twice in 1 ml 1x PBS (130 mM NaCl, 10 mM Na_2HPO_4, 10 mM NaH_2PO_4, pH 7.2), resuspended in 1x PBS containing 10µg 4',6'-diamidino-2-phenylindole (DAPI) ml^{-1} (final concentration), and incubated for at least 10 min at room temperature protected from light. Cells were then pelleted and washed twice with 1x PBS, and observed using a fluorescence microscope with the proper UV filter set.

4.3.4 Comparative analysis of cell size

Analysis of cell size was carried out by flow cytometry, as described by Pina-Vaz and co-workers (2001), with some adaptations. Briefly, cells were harvested in stationary phase (O.D.~4.0) after growing overnight in MM at 30ºC with 180 rpm. After centrifugation, cells were washed in PBS buffer (pH 7.0) and resuspended in 1 ml of the same buffer supplemented with 2% (w/v) glucose. Cells were then sonicated, counted in a Newbauer chamber, diluted to a density of $1x10^6$ cells/ml and analysed by flow cytometry (Pina-Vaz *et al.*, 2001).

4.3.5 Contour-clamped homogeneous electric field (CHEF) electrophoresis

Preparation of cells for CHEF analysis was performed as described previously by Selmecki and colleagues (2008). Briefly, cells were grown overnight in liquid YPD, gently collected by centrifugation, washed several times with 50 mM EDTA, re-suspended in 1% low melt agarose and treated with lyticase and with a Proteinase K solution. Whole chromosome separation was performed on a BioRad CHEF-DR III with the following program: 60- to 120-s switch, 6.0 V.cm-1, 120ºangle, for 36 hours, followed by a 120- to 300-s switch, 4.5 V.cm-1, 120ºangle, for 12 hours (Selmecki *et al.*, 2008).

4.3.6 Multiplex PCR assay

Detection of cases of chromosomal aneuploidy, formation of isochromosomes or loss of chromosome ends was carried out as described by Arbour and co-workers (2009). The genomic DNA was purified using the DNA purification kit (Promega) according to the manufacturer's instructions. PCR reactions (total volume of 50 mL) contained 1x DreamTaq buffer, 0.2 µM equimolar primers

mixture (set A and B), 0.2 mM dNTP mix, 0.375 units of DreamTaq Polymerase (Fermentas) and 50 ng of purified genomic DNA. Thermal cycling was carried out in a BioRad thermocycler with a denaturation step of 95ºC for 3 minutes, 30 cycles with 30 seconds denaturation at 95ºC, 30 seconds annealing at 52ºC, 1 minute elongation at 72ºC and the last elongation step at 72 ºC for 5 minutes. Following amplification, 1 µL of the PCR reaction was loaded into the well of a Bioanalyzer chip prepared according to the manufacturer's protocol for the DNA 7500 Lab Chips (Agilent Technologies). The aneuploidy of the mutant DNA was determined by the relative ratio of the peak height of the mutant and wild-type (SC5314) DNA fragments in the chromatogram. The ratio of each chromosome was then divided by the median of the ratio of all chromosomes (Arbour *et al.*, 2009). At least 3 replicate experiments on several individual colonies were used for identification of aneuploidies.

4.3.7 DNA extraction

Genomic DNA extraction was carried out using the Genomic-tip 100/G kit (Qiagen) according to the manufacturer's protocol. Quantification and quality assessment were performed using the Picogreen fluorescence based quantification assay.

4.3.8 Illumina sequencing, alignment and analysis of sequences

For Illumina sequencing, genomic DNA was prepared and sequenced using manufacturer-supplied protocols and reagents, as follows. One library per sample was constructed using Illumina DNA Sample Prep standard protocol and with an insert size of 400-500 bp. Briefly, 5 µg of high molecular weight genomic DNA (gDNA) was fragmented by Covaris sonication device. Following sonication, DNA

fragments were end-repaired and A-tailed. Adapters were then ligated via a 3′ thymine overhang. Finally, ligated fragments were amplified by PCR. The library was applied to an Illumina flowcell for cluster generation. Sequencing was performed on a Genome Analyzer IIx instrument using ~150 bp paired-end reads.

Raw sequence data, 146 bp paired end reads with expected insert size of 400-500 bp, from each sample were trimmed by removing consecutive bases on both 5'- and 3' flanks with base quality less than 20. Trimmed reads that did not pass filtering criteria for ambiguity (N content < 5%), complexity (score >= 10), length (50 bases or longer), and average base quality >=20 were removed using bamtools (Barnett *et al.*, 2011) (Annex G1).

Remaining reads were mapped to the reference genome of *Candida albicans*, obtained from the Candida genome database (http://www.candidagenome.org/, assembly 21) using BWA, version 0.5.9 (Li and Durbin, 2010). Processing and filtering of mapped reads were done using Samtools, version 0.1.17 (Li *et al.*, 2009). After removal of duplicates, read pairs aligning to opposite strands, or where predicted insert size did not match actual size, were removed. Additionally, read pairs were removed where one or both reads had low mapping quality (MQ<20) or had less than 95% sequence identity to the reference (Annex G1). Over all samples, about 49.4 % of the read pairs remained after this filtering, covering on average 99.4 % of the genome length for an average alignment depth of 56.8 to 96.2 reads per base (Annex G2).

Mapped reads were analysed using Samtools. The control T0 strain used in this experiment is genetically slightly different from the SC5314, which has been used for the construction of the reference genome. We identified genomic regions not well covered by the reference assembly and removed reads mapping to these. These regions were identified by reads covering genomic positions for which three or more allele types had been mapped with good quality, as well as reads covering three or more bi-allelic genomic positions within 10 base pairs distance, where the least frequent allele type was represented by 20 % or more of the reads for all three of the positions (Annex G3). On average, 0.4 % of the length of the genome and 3.3 % of the total number of remaining reads were removed in this way (Annex G1).

Samtools was used to produce read pileups, detect single nucleotide variants and call genotypes. Small insertions and deletions (indels) were not called. Bases with low base quality or with read depth less than 3 or higher than twice the sample average coverage were called as unknown genotype. For the four samples containing the $tRNA_{CAG}^{Leu}$ gene, we analysed the number and genomic class, such as coding, non-coding or repeat region, of bases with different genotype compared to the control strain. All samples showed allele frequency spectra indicative of predominantely diploid characteristics with relatively low amount of possible aneuploidy (Annex G4, G5). SNPs in coding regions were also analysed for changes in codon usage compared to the control strain. The same comparisons were done between strains T1 and T2, and between T2 and its derived strains T2KO1 and T2KO2.

4.4 Results

FACS analysis of cells misincorporating leucine at CUG positions. Ambiguous and reverted strains were grown in YPD medium at 30ºC and several isolates from each strain were analysed by flow cytometry to determine genome size changes (ploidy changes) (Figure 4-1). *C. albicans* exists as a diploid (2N) yeast, however some clones often become polyploid (Scherer and Magee, 1990). In our study, flow cytometric analysis confirmed that each mistranslating strain was diploid (2N), or close to diploid, in DNA content, as judged from DNA staining with an intercalating fluorescent dye (SYBR green I). This was confirmed by nuclear observation using epifluorescence microscopy. Mistranslating T0, T1, T2, T0KO1, T1KO1 and T2KO1 cells and reverted T2KO2 cells were stained with DAPI and showed no nuclear morphological alterations and only one nucleus per cell (Figure 4-2). Therefore, reversion of CUG identity did not increase yeast ploidy conversely to what has been previously described (Miranda *et al*., 2007).

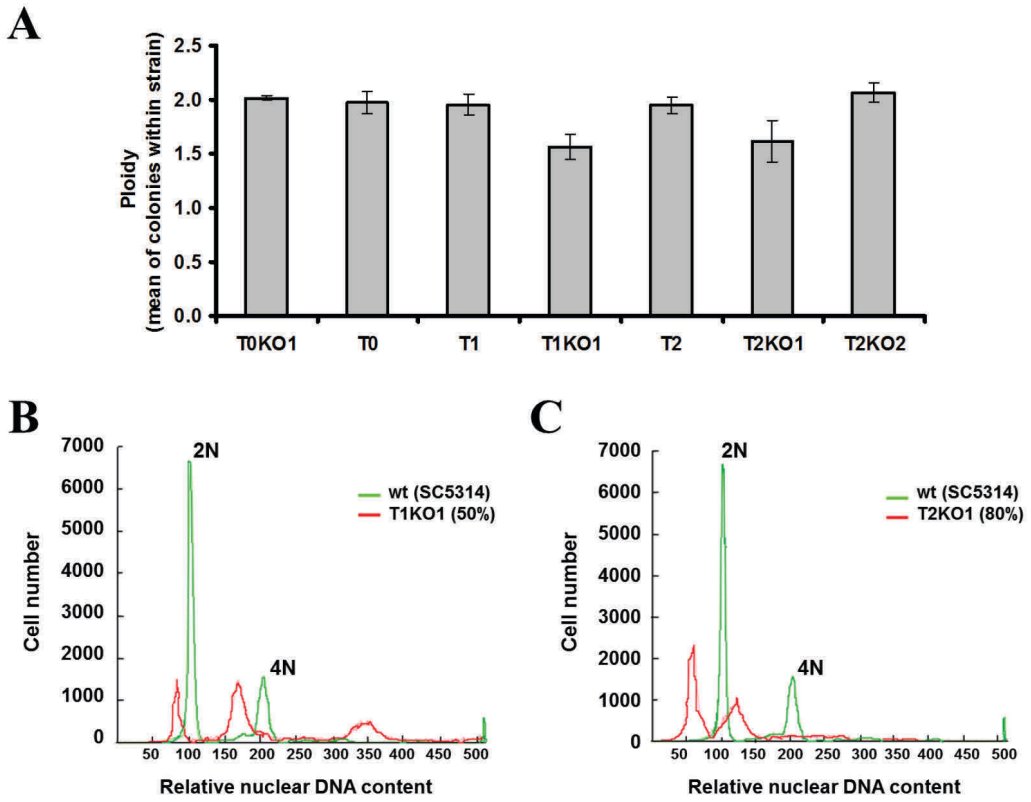


Figure 4-1. Nuclear DNA content of highly ambiguous and reverted strains. Nuclear DNA content was measured by flow cytometry in control and mistranslating strains. **A.** The majority of the strains showed no alteration in nuclear DNA content (2N); exceptions were T1KO1 and T2KO1 which were less than diploid. **B.** Representative histogram of the strain T1KO1 whose clones were ~1.56N. **C.** Representative histogram of the strain T2KO1 whose clones were ~1.61N.

However, subtle differences between strains were observed in the DNA profiles obtained by flow citometry. For example, distinct peaks were evident, representing non-replicated (G1 phase) and replicated (G2 phase) DNA. Also, many of the FACS profiles had a larger peak in addition to the G1 and G2 peaks, as seen in Figure 4-1B at ~350. This peak usually represents cells that have undergone mitosis, but have not separated and passed through the cytometer parallel to the laser so that 2 - 2N nuclei are detected as 1 - 4N. In some mistranslating strains the majority of the cells contained non-replicated DNA (G1) as in the control SC5314 (Figure 4-1C), while others had an almost equal

distribution of cells with unreplicated (G1) and replicated DNA (G2) (Figure 4-1B). This was highly variable depending on the different colonies of the same strain that were analysed. The calculated ploidy for each different clone of the same strain was averaged and most strains had a nuclear DNA content of ~2N (Figure 4-1A). Surprisingly, two strains had clones that were less than 2N. Strain T1KO1 showed a mean ploidy of 1.56N, which represents a 20% decrease from diploid (Figure 4-1A and Figure 4-1B). Strain T2KO1 showed a mean ploidy of 1.61N, which also represents a 20% decrease from diploid (Figure 4-1A and Figure 4-1C). Although we did not have evidences for pure haploid cells, there was a greater tendency for being less than diploid rather than >2N, suggesting that CUG ambiguity altered genome organization.

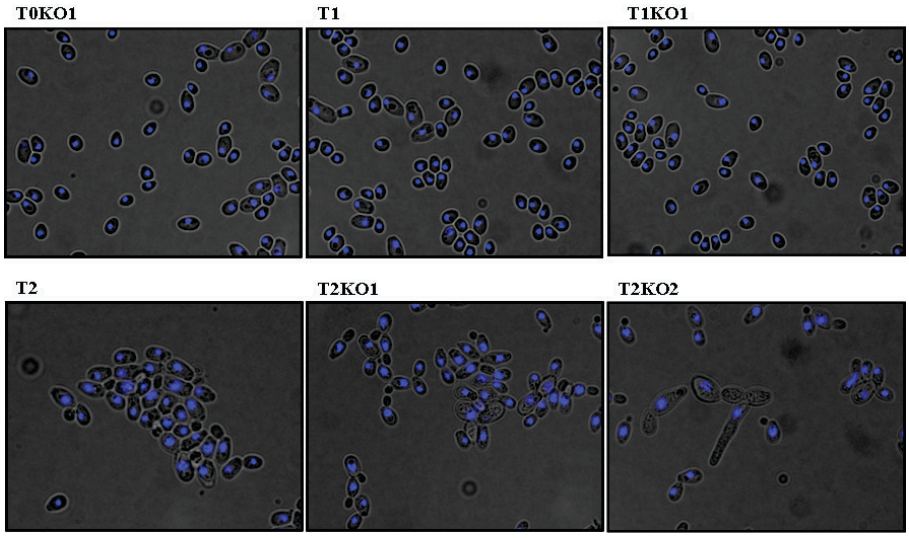


Figure 4-2. Nuclear morphologies of highly ambiguous and reverted strains. DAPI staining showing that mistranslating cells have a single nucleus. Nuclear aberrations or multiple nuclei were not observed.

Flow cytometry also showed that cell size and shape, interpreted from forward and side scatter plots, were highly variable (Figure 4-3), however no obvious correlation between cell ploidy and cell complexity could be established. The forward scatter measurement depicts cell size and the side scatter measurement provides information about organelles or budding chains of cells.

The complexity of the strains increased with higher levels of mistranslation and strain T1KO1 (50% of leucine misincorporation) showed the highest level of complexity.

Although flow cytometry was useful to detect whole-genome ploidy shifts among different isolates of mistranslating cells, it could not determine other types of genome variability events, such as translocation, deletion, amplification, loss of particular chromosomes and chromosome length polymorphisms (Selmecki *et al.*, 2010).

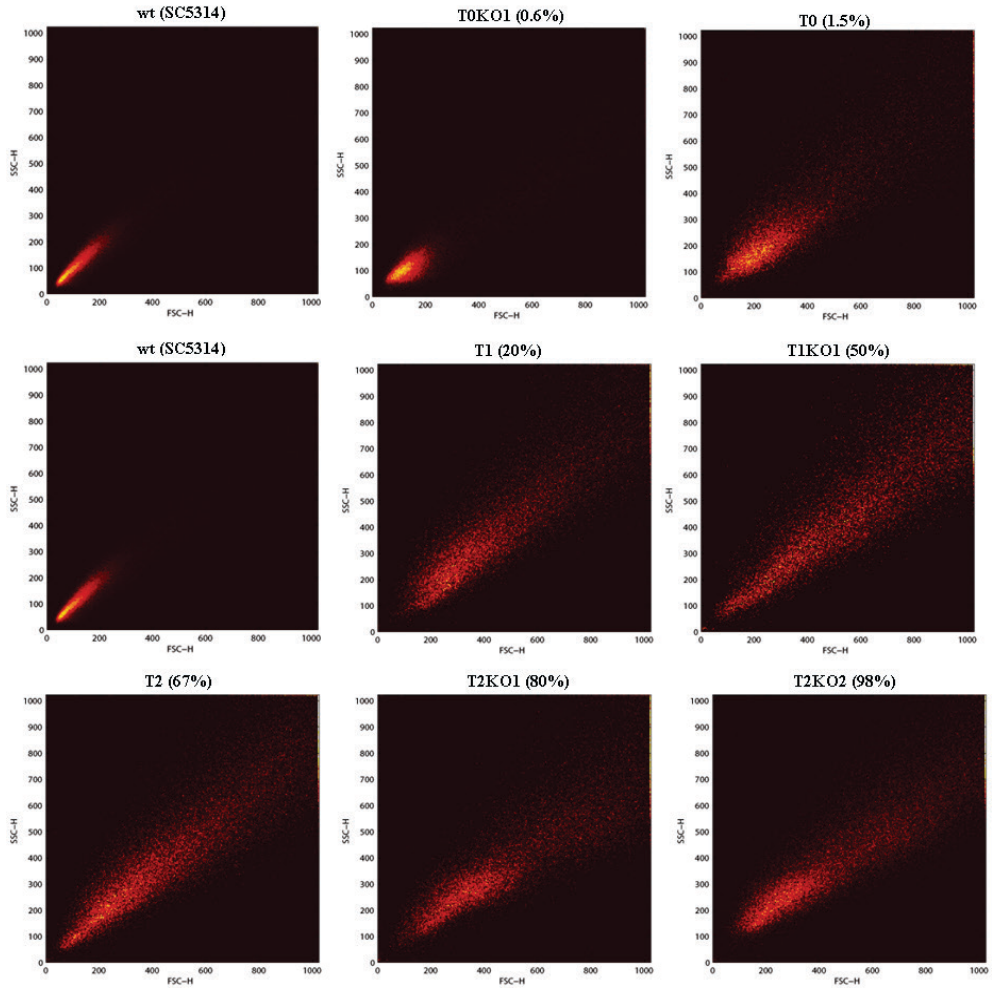


Figure 4-3. Analysis of cell size and complexity. Analysis of cell size was carried out by flow cytometry. Picture represents a collection of dot plot graphs in which the X-axis is a measure of forward scatter of light (FSH) and the Y-axis is a measure of side scatter of light (SSH). The forward scatter measurement depicts cell size and the side scatter measurement provides information about organelles or budding chains of cells. The complexity of the strains increased with higher levels of mistranslation. Strain T1KO1 (50% of leucine misincorporation) showed the highest level of complexity.

CHEF analysis of the *C. albicans* karyotype. CHEF (contour-clamped homogenous electrophoretic field) gel analysis was used to identify chromosome alterations. In standard karyotype analysis, the number of *C. albicans* chromosome bands is 8, hence the diploid chromosome number is 16, but chromosomes contain largely homologous genetic information that migrate together on pulse-field gels (Magee *et al.*, 1992). In our experiment, the pulsed-field gel electrophoresis separated the 8 *C. albicans* chromosomes and it was optimized to resolve smaller chromosomes. Karyotyping revealed no obvious differences in chromosome copy number between the mistranslating strains and the control SC5314 strain. But, strains with high level of leucine misincorporation at CUGs (T2KO1 and T2KO2) showed slight changes in the size of small chromosomes (Figure 4-4A). For example, the highly ambiguous strain T2KO1 had slight differences in chromosomes 5, 6 and 7 sizes (Figure 4-4B), supporting the idea that ambiguous CUG decoding induced genome destabilization.

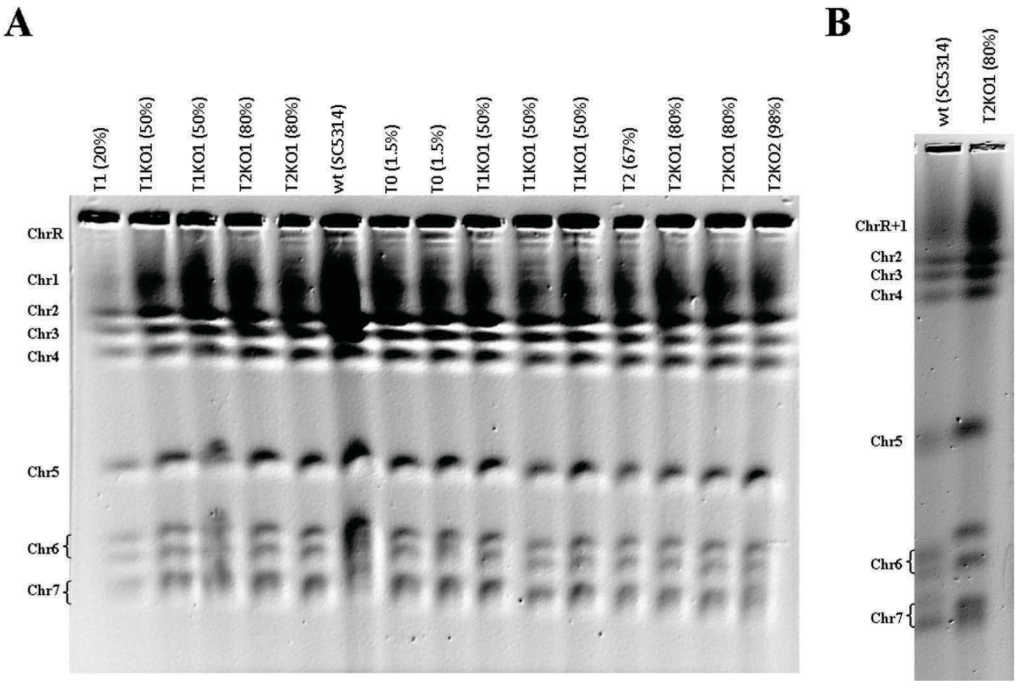


Figure 4-4. Contour-clamped homogeneous electric field (CHEF) electrophoresis of mistranslating strains. A. CHEF analysis showed no obvious differences in chromosome copy number and no major changes in chromosome sizes, except in smaller chromosomes where slight changes were observed. **B.** Strain T2KO1 had slight differences in chromosomes 5, 6 and 7 sizes.

Whole genome re-sequencing engineered *C. albicans* strains. The above data suggested that CUG ambiguity induces subtle genome rearrangements. In order to test this hypothesis, we have re-sequenced the whole genome of strains T0, T1, T2, T2KO1 and T2KO2 in collaboration with partners from the European Sybarys project.

Illumina genomic DNA libraries were generated from DNA isolated from the control strain T0 and four mistranslating strains, namely T1, T2, T2KO1 and T2KO2. We collected between 6,141,944 and 11,730,542 paired-end, 150 bp reads from the T0 and T2KO2 genomes, respectively, on an Illumina Genome Analyzer II platform. Reads that did not pass filtering criteria for ambiguity (N content < 5%), complexity (score ≥ 10), length (50 bases or longer), and average base quality ≥20 were removed using bamtools (Table 4-2 and annex G1) (Barnett *et al.*, 2011). Remaining reads were mapped to the reference genome of *Candida albicans* SC5314, obtained from the Candida genome database (http://www.candidagenome.org/, assembly 21) using BWA, version 0.5.9 (Li and Durbin, 2010). Following quality filtering, these reads yielded a coverage of ~99.4% of the genome length for an average alignment depth between 56.8 to 96.2 reads per base for T0 and T2, respectively (Table 4-2 and annex G2).

Table 4-2. Read mapping statistics, average alignment depth and reference genome coverage of mapping.

	T0	T1	T2	T2KO1	T2KO2
Raw reads	6,141,944	8,437,847	10,336,605	6,923,919	11,730,542
Reads ready to align [1)]	5,125,425	6,975,400	8,552,088	5,708,627	6,012,173
Average alignment depth	56.8	76.4	96.2	62.7	91.0
Genome coverage (%)	99.4	99.4	99.4	99.4	99.6

[1)] after filtering reads for ambiguity

By analysing the mapped reads, we were able to see that the control T0 strain used in this experiment is genetically slightly different from the SC5314 strain, which has been used for the construction of the reference genome. One large fraction of the reads had perfect sequence identity with the reference regions, whereas another small portion had frequent differences, indicating a different sequence source not present in the reference genome sequence (Annex G3). These genomic regions not well covered by the reference assembly were filtered out, as described in the methods section. On average, 0.4 % of the length of the genome and 3.3 % of the total number of remaining reads were removed this way. After this filtering step, Samtools was used to produce read pileups, detect single nucleotide variants and call genotypes.

Genome sequence data analysis showed that all samples had allele frequency spectra indicative of predominantly diploid characteristics with relatively low amount of putative aneuploidy (Annex G4, G5). This result is in line with the above mentioned flow cytometry results which showed that mistranslating strains are mainly diploid (or close to diploid). Furthermore, read-depth information from the high-throughput sequencing data was relatively uniform across the different genomes (Annex G5), suggesting that copy number variations between strains is low or non-existing.

To identify single nucleotide polymorphisms (SNPs), we have examined base calls using criteria that permitted comparing nucleotide positions in the mistranslating and reference short read data sets over the majority of the genome. We performed SNP calling following an approach that monitors commonly applied filters for read-depth and quality score. Thus, bases with low base quality or with read depth less than 3 or higher than twice the sample average coverage were discarded. Applying these criteria, we detected between 45,297 and 55,429 SNPs from a unique set of 61,312 locations in the five sequenced strains. These values represent an average rate of 1 SNP per 315 bases in strain T2KO1 and 1 SNP per 258 bases in strain T1 (Table 4-3). Therefore, the average rate of SNPs across different mistranslating genomes is not much different. Furthermore, these

numbers are consistent with the previously estimated >48,000 SNPs in strain SC5314 and >41,000 SNPs in strain WO1, giving 1 SNP per 330 and 390 bases, respectively (Jones *et al.*, 2004;Braun *et al.*, 2005;Butler *et al.*, 2009).

Table 4-3. Single nucleotide polymorphisms in each sample.

Chromosome	T0	T1	T2	T2KO1	T2KO2
chr1	15,428	15,561	15,526	15,482	15,820
chr2	4,665	4,757	4,800	4,719	4,908
chr3	3,679	3,670	3,714	3,681	3,784
chr4	8,145	8,198	8,190	8,122	8,376
chr5	6,144	6,234	6,229	4,507	4,666
chr6	6,651	6,761	6,772	6,714	7,010
chr7	1,583	1,630	1,621	1,639	1,679
chrR	8,431	8,527	8,577	433	8,824
all	54,726	55,338	55,429	45,297	55,067
Avg. rate of SNPs in NT genome	261	258	258	315	259

Since the T0 strain was the wild type ambiguous strain used to construct all the high ambiguous strains, we also analysed the number and type of SNPs for the four strains containing the $tRNA_{CAG}^{Leu}$ (T1, T2, T2KO1 and T2KO2) in comparison with the control strain T0. When compared to T0, a near-complete loss of heterozygosity (LOH) in a 300 kb region on chromosome 5 (825,000 - 1,136,000) was observed in strains T2KO1 and T2KO2, and for the entire chromosome R in the T2KO1 strain (Figure 4-5 A-B and annex G6-7). The affected region on chromosome 5 contains 178 ORFs but most of those genes have an unknown function. In an effort to characterize the LOH identified in T2KO1 and T2KO2, Gene Ontology (GO) process categorization was carried out (Figure 4-5C). Among the genes that are characterized, the majority is involved in the

regulation of biological processes (21.3%) and organelle organization (13.5%). Moreover, 9.6% of the affected genes have functions related to the response to stress, 6.2% are specifically related to the response to drugs, 9% are involved in filamentous growth, 2.8% are involved in pathogenesis and 2.2% are related to conjugation (Figure 4-5C). Some of the genes within these categories are *FAV2* and *CAG1* that are part of a functional mating response pathway (Dignard and Whiteway, 2006), gene *GLN3* that encodes a GATA transcription factor involved in the regulation of filamentous growth induced by nitrogen starvation (Dabas and Morschhauser, 2007) and *RIM13* that encodes a protease involved in the pH response pathway (Davis, 2003).

As mentioned above, 6.2% of the genes affected by LOH on chromosome 5 are specifically related to the response to drugs. These genes included *orf19.3218* that encodes a putative multidrug resistance protein (Doedt *et al.*, 2004). Other LOH affected genes were *CCT3*, *TIP1*, *orf19.3938*, *orf19.3936* and *orf19.3894* whose deletion confer hypersensitivity to a toxic ergosterol (Xu *et al*., 2007), *PGA4* which encodes a GPI-anchored cell surface protein (Sorgo *et al*., 2011), *CNB1* that codifies the regulatory subunit of calcineurin B (Ca^{2+}-calmodulin-regulated S/T protein phosphatase) (Cruz *et al.*, 2002;Singh *et al.*, 2009), *MID1* that encodes a calcium ion channel (Karababa *et al*., 2004) and *YUH2* which encodes a sumoylation target (Leach *et al.*, 2011). Although LOH on chromosome 5 and drug resistance have been previously associated (Selmecki *et al.*, 2006;Butler *et al.*, 2009;Selmecki *et al.*, 2009), the location of the LOH events in our study is different from the previous reports. For example, homozygosis of *TAC1* is frequently associated with increased azole resistance due to the fact that *TAC1* is required for the upregulation of the ABC-transporter genes *CDR1* and *CDR2* (Coste *et al*., 2006). LOH events encompassing the drug target *ERG11* have also been shown to contribute directly to the acquisition of high level azole resistance (Selmecki *et al.*, 2006). Both cases concern the left arm of chromosome 5 while in our study LOH occurred in the right arm.

Nevertheless, considering the genes that were affected by LOH on chromosome 5, they can reveal genetic variability associated with mating, growth

and stress resistance by exposing the phenotypes associated with recessive alleles.

As for LOH on the entire chromosome R in T2KO1, it includes LOH for 989 ORFs, namely the ribosomal DNA (rDNA) cistrons. The presence of these tandemly organized units in chromosome R generates chromosome instability which may explain the sensitivity of this chromosome to the stress induced by mistranslation. In fact, the number of rRNA units varies from 9 to 180 in different strains (Rustchenko, 2007). One intriguing observation of the LOH on chromosome R was that strain T2KO2 did not show it. Considering that T2KO2 is derived from T2KO1, one was expecting that the reverted strain also had LOH on chromosome R. This may be explained by the fact that only one isolate of each strain was sequenced, although multiple isolates were used for strain construction. Consequently, the T2KO1 isolate that was sequenced may not have been the one that originated the T2KO2 isolate. Since chromosome R is highly unstable, one should sequence multiple isolates of the same strain in future studies.

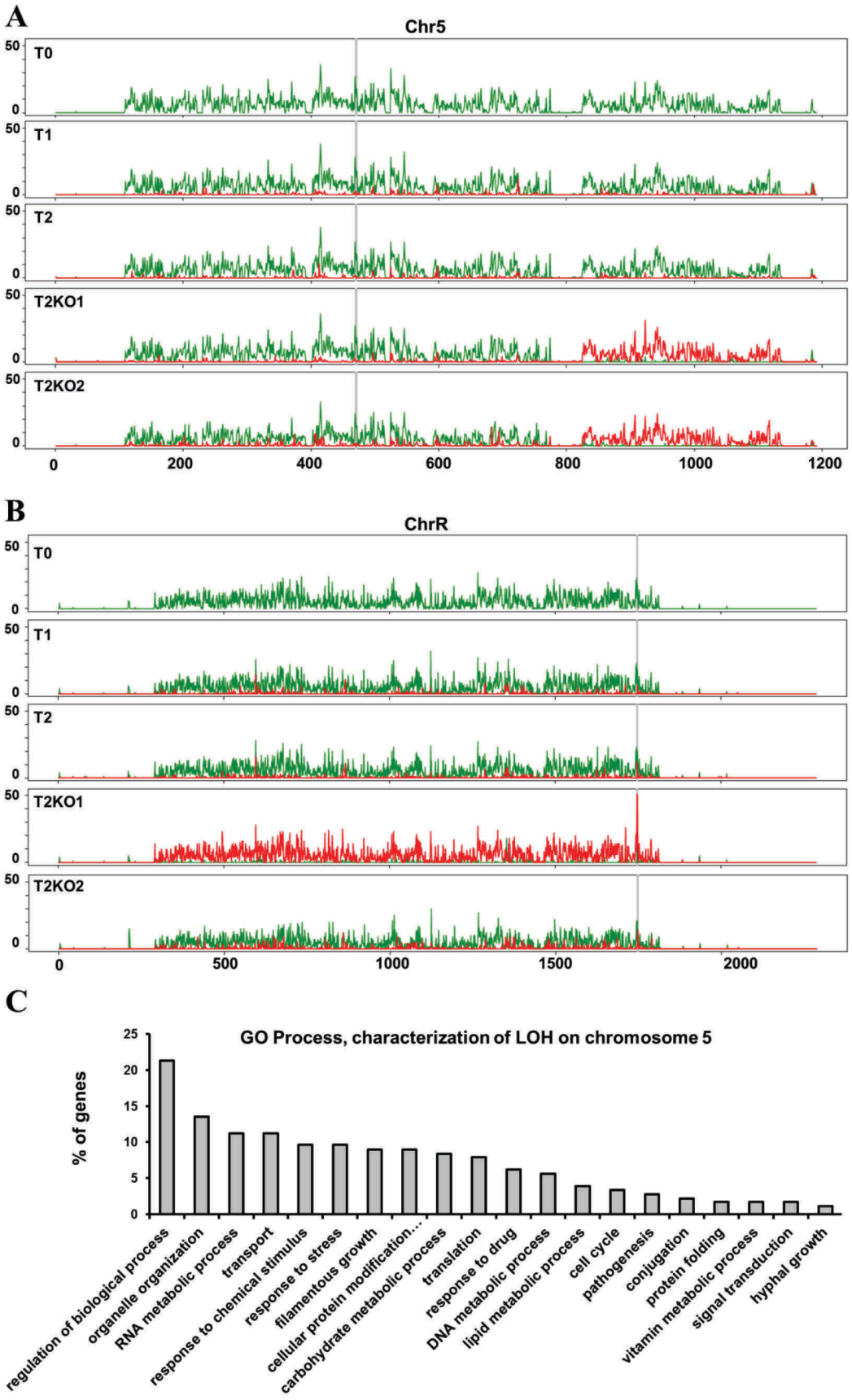

A
Chr5
T0
T1
T2
T2KO1
T2KO2
50
0
0
200
400
600
800
1000
1200
B
ChrR
T0
T1
T2
T2KO1
T2KO2
50
0
0
500
1000
1500
2000
C
GO Process, characterization of LOH on chromosome 5
% of genes
25
20
15
10
5
0
regulation of biological process
organelle organization
RNA metabolic process
transport
response to chemical stimulus
response to stress
filamentous growth
cellular protein modification...
carbohydrate metabolic process
translation
response to drug
DNA metabolic process
lipid metabolic process
cell cycle
pathogenesis
conjugation
protein folding
vitamin metabolic process
signal transduction
hyphal growth

Figure 4-5. The SNP density in the engineered *C. albicans* strains. A and B. The SNP density (SNPs/kilobase, shaded green) varies along the chromosomes (shown as subfigures) but are essentially similar between the five strains (represented by tracks within each subfigure). There are homozygous stretches matching in all samples, as well. The density of SNPs with loss of heterozygosity (LOH) in a derived strain compared to T0 are plotted in red, showing a region on Chromosome 5 in both T2KO1 and T2KO2, and the entire Chromosome R in T2KO1, where all heterozygote SNPs have changed to homozygous. Grey vertical line indicates the centromere. **C.** Gene ontology (GO) process terms for the 178 ORFs affected by LOH on chromosome 5. The y-axis displays the percentage of genes within each category from the total of 178 affected by LOH. GO process characterization was performed using the *Candida* Genome Database.

LOH events are common in *C.albicans* and can arise through several mechanisms. For example, in our data the extent of LOH was different between chromosome 5 and chromosome R, while chromosome 5 presented a short LOH tract (region 825,000 - 1,136,000), chromosome R showed a whole-chromosome (whole-chr) LOH. The first is usually originated through gene conversion or double crossovers and the second is frequently a product of chromosome nondisjunction that occurs because of defects in centromere/kinetochore/mitotic spindle function (Selmecki *et al.*, 2010). Whole-chr LOH usually results in aneuploidy, an imbalance in the number of chromosomes, which may have happened in the chromosome R of the sequenced T2KO1 isolate. To confirm this hypothesis and test the possibility of variability among different isolates of the same strain, we performed a quantitative multiplex PCR assay on numerous isolates of strains T2KO1 and T2KO2 (Arbour *et al.*, 2009). This test was developed by Arbour and co-workers (2009) and can easily detect cases of chromosomal aneuploidy. Each assay consists of eight pairs of primers that are specific for unique regions near the left or the right arm of each chromosome. As expected, the PCR reaction produced eight amplicons of different sizes, one from each chromosome (Figure 4-6A). These assays were conducted using the primer set for the right chromosomes arms (Arbour *et al.*, 2009) with genomic DNA extracted from control

T0 and mistranslating strains. Quantification of the corresponding amplicons was carried out using an Agilent Bioanalyzer, a precise capillary electrophoresis that produces reproducible results and allows one to directly compare the elution profiles. The quantification of each amplicon was normalized with the corresponding amplicon of the control T0 (Figure 4-6 B-E). The aneuploidy on chromosome R in isolates 1 from T2KO1 and T2KO2 was clearly visible (Figure 4-6B and Figure 4-6E). Conversely, isolates 2 did not show the same aneuploidy on chromosome R, which proves the variability among different isolates of the same strain. Also, no other chromosome besides chromosome R showed evidences of aneuploidy.

When we compared loss of heterozygosity (LOH) between T2KO1 and T2KO2, 2,195 LOH loci were detected in both T2KO1 and T2KO2. Among these, 1,699 SNPs converted to opposite homozygotes in the two knock-out strains. This surprising result was particularly evident in chromosome 5 where 1,652 SNPs (approximately 75% of the total LOH loci) showed different homozygotes (Figure 4-7). The remaining 47 SNPs that converted to opposite homozygotes in the two knock-out strains were located in chromosome R. Sanger sequencing validation of genotypes for 15 SNPs in T2KO1 and T2KO2 confirmed that the SNPs in the LOH region on chromosome 5 have converted to the opposite homozygotes in the two knockout strains (Annex G8). This occurrence may be explained by the LOH caused by mitotic recombination, where a region with “AB” genotype SNPs, after replication and crossover, yield one daughter cell with “AA” genotype and one with “BB”, inducing a 50/50 population with either of the two homozygotes for SNPs in the LOH region. When plating and sampling a single colony for sequencing, we will only see one of the two homozygous chromosome variants.

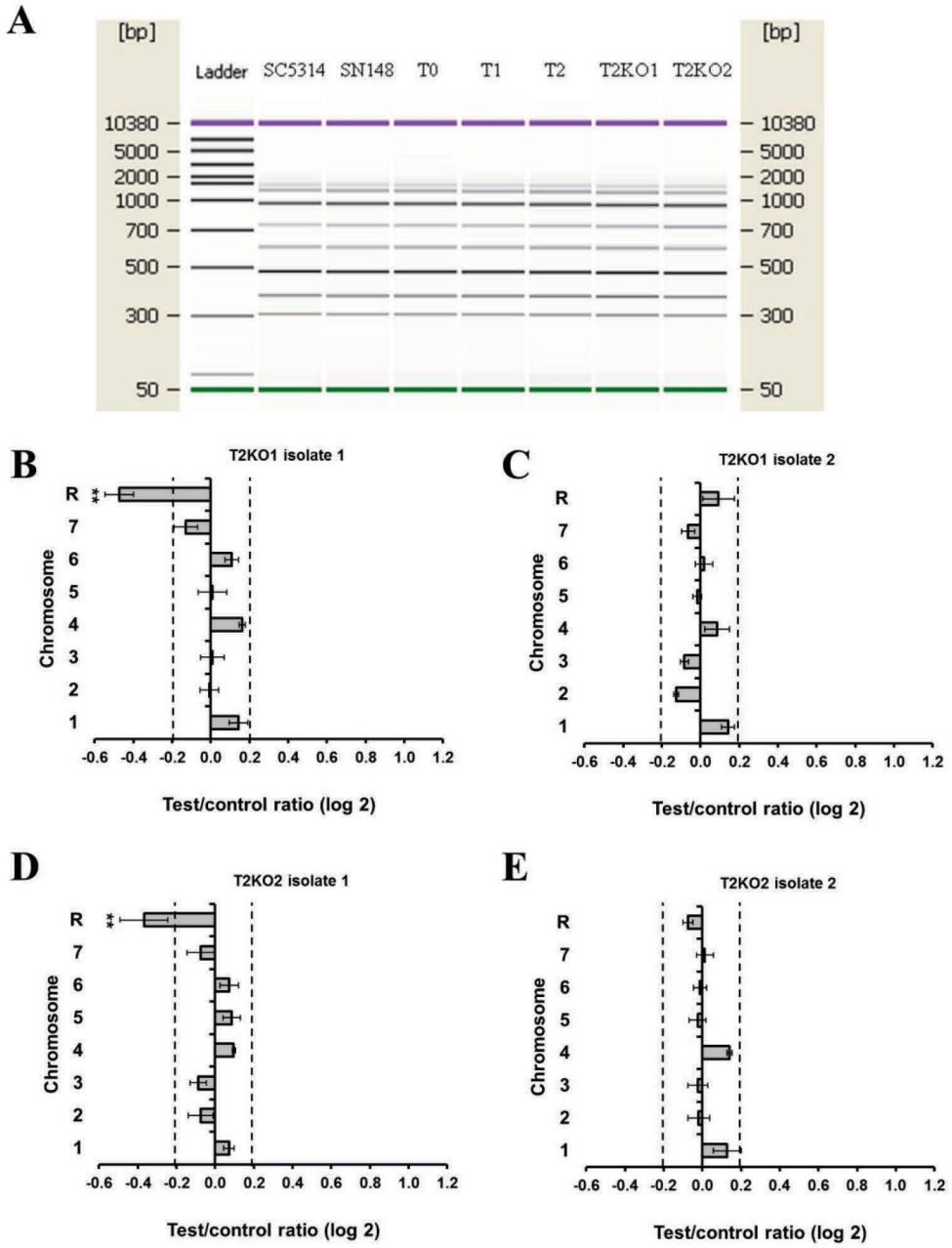


Figure 4-6. Aneuploidy analysis using a multiplex PCR assay. A. Bioanalyzer profiles of multiplex PCR reactions using primer set B (right arm). Elution profiles allowed the identification of amplicons with different abundance. **B-E.** The assay was conducted on genomic DNA from strains T0 (control strain) and mistranslating strains. Graphs represent the mistranslating/T0 ratio of the median normalized abundance of amplicon on the x-axis and each chromosome on the y-axis. A log2 ratio above 0.2 was considered to be significant and indicative of aneuploidy for these chromosomes. Data represent the mean ± standard deviation of 3 independent experiments (**p<0.01 one-way Anova post Dunnett's comparison test with CI of 95%, relative to control T0).

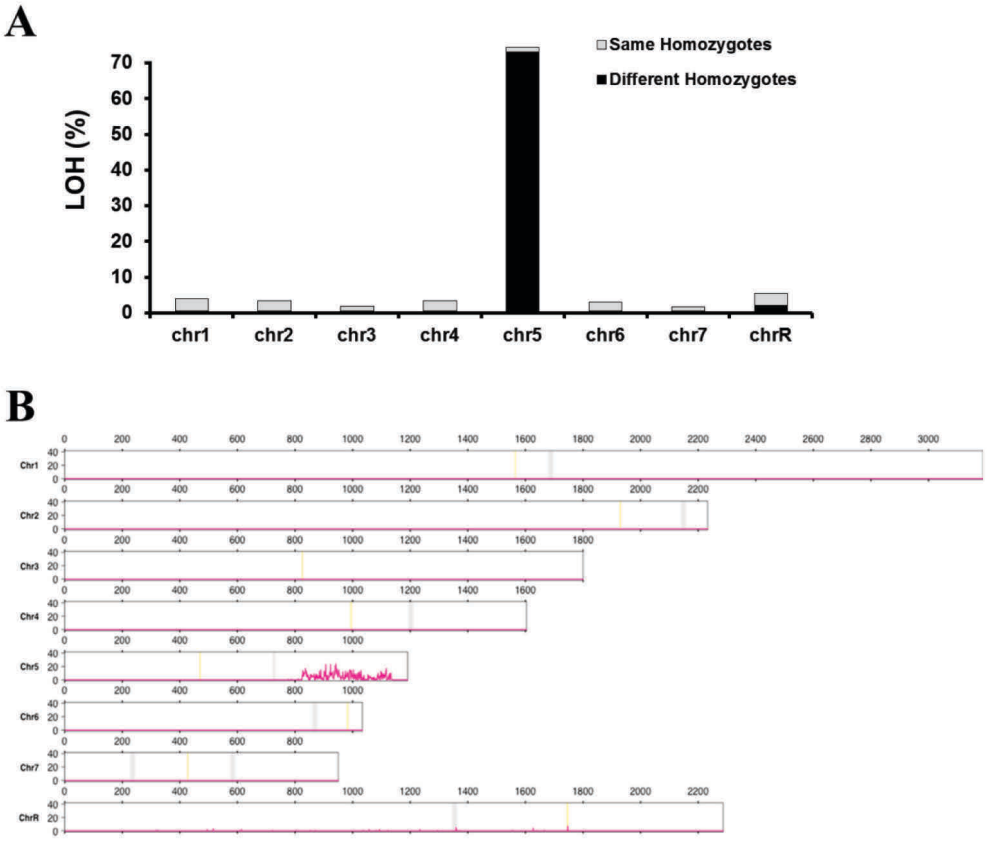


Figure 4-7. Comparison of loss of heterozygosity (LOH) between T2KO1 and T2KO2. **A.** For 2,195 LOH loci in both T2KO1 and T2KO2, only two chromosomes showed different homozygotes: 1,652 loci in chromosome 5 and 47 loci in chromosome R. **B.** Chromosome positions with different homozygotes between two knock-out samples is plotted for windows of 1 kilobase. In T2KO1 and T2KO2, the SNPs that lose heterozygosity go to different homozygote genotypes in the LOH region on chromosome 5.

As mentioned before, we detected between 45,297 and 55,429 SNPs from a unique set of 61,312 locations in the five sequenced strains. Despite this, the average rate of SNPs was not much different among the different strains with increasing leucine misincorporation. However, we wondered if the LOH regions of chromosomes 5 and R in the high mistranslating strains were interfering with the analysis. Therefore, a subsequent SNPs calling was performed but the two LOH

regions were excluded from the analysis, namely the 825,000~1,136,000 region of chromosome 5 and the 291,000~1,814,000 region of chromosome R. Out of 49,922 SNPs observed outside the LOH regions, each strain had heterozygote SNPs at between 44,673 and 46,256 of these positions. Again, the five strains have essentially an unchanging number of SNPs, when discounting the LOH regions on chromosome 5 and R (Figure 4-8A) and only T2KO1, which is affected by LOH on the entire chromosome R, seems to be different. Interestingly, in the part of the genome not affected by LOH, we have observed an increasing number of unique SNPs in strains with increasing leucine misincorporation (Figure 4-8B). This effect is particularly evident in the reverted strain which has 692 unique SNPs not found in any other strain, nearly three times as many as the other strains (Figure 4-8B).

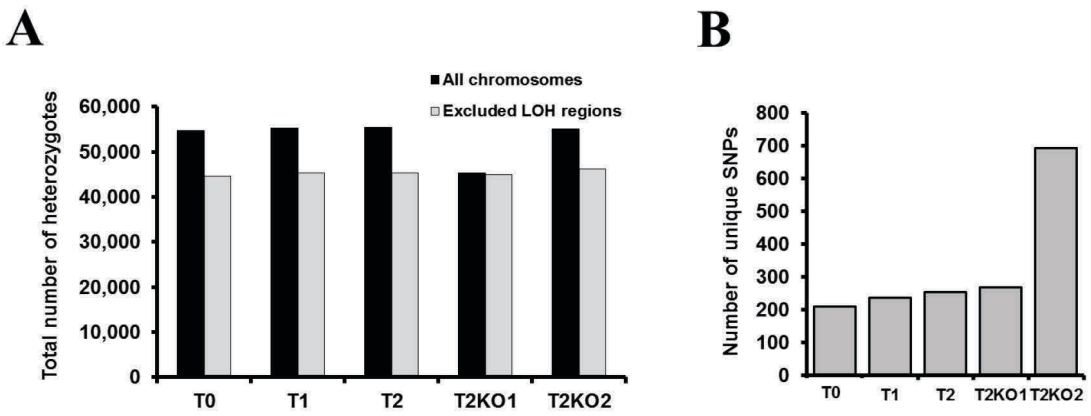


Figure 4-8. Single nucleotide polymorphisms analysis excluding LOH regions. A. The five strains have similar number of SNPs, when discounting the LOH regions on chromosomes 5 and R (grey bars). When the entire genomes were considered (black bars), T2KO1 was heavily affected by LOH on the entire chromosome R. **B.** Changes outside LOH regions are located at an increasing number of unique heterozygote SNPs as the degree of leucine misincorporation at CUGs increased.

LOH regions were also excluded in a genotype calling (SNPs calling) analysis using T0 as the reference genome. Again, strains showed an increasing number of genotype changes at polymorphic locations with increasing degree of

mistranslation. For example, T2KO2 changes genotype compared to control T0 at 2,879 SNPs, nearly 600 more than the other three mistranslating strains (Figure 4-9A and Annex G9). Also, genotype change analysis was heavily affected by the inclusion of the LOH regions, particularly in the case of strain T2KO1 (Figure 4-9A).

The increased rate of genotype change from control is accumulated mostly in the non-coding region of the DNA. For example, strain T2KO2 had 1924 genotype changes located in non-coding positions while only 955 were at coding positions (Figure 4-9B). This trend was common to all strains and the rate of genotype change in coding regions was similar for all strains (between 31.9% and 34.6% of the SNPs in T1 and T2KO1, respectively) (Annex G9). In *C. albicans*, 63.0% of the genome constitutes coding regions and the lower proportion of SNPs in coding regions expected by a uniform random rate can be explained by the increased risk for a functionally disruptive mutation in those regions. A mutation in a non-coding region is much less likely to affect the organism's phenotype than one in a coding region. Hence, it is more likely for these non-coding regions to accumulate mutations.

Analogously, there was no shift in nonsynonymous mutation rate in any of the strains (between 13.6% and 14.1% of the SNPs) (Annex G9). For instance, strains T1, T2, T2KO1 and T2KO2 had 312, 297, 314 and 399 nonsynonymous mutations, respectively (Figure 4-9C). These nonsynonymous changes were located in numerous genes among all strains but most of these genes were common between all strains. Remarkably, 86 of those genes had genotype changes in the reverted strain only (Annex G10). Of the 399 nonsynonymous mutations of T2KO2, 131 were located at these specific genes. The list of these 86 genes provided in Annex G10 includes *TIF4631* gene that encodes the translation initiation factor eIF4G, *DDC1* that encodes a DNA damage checkpoint protein, *TSC2* gene encoding a GTPase-activating protein involved in the control of filamentous growth, *SGE1* that encodes a multidrug resistance factor, *ATM1* that encodes a member of the multidrug resistance (*MDR*) subfamily of proteins and members of the *ALS* protein family that have an important role in adhesion and germ tube induction. Functional class analysis of this set of genes showed that

nonsynonymous mutations in T2KO2 accumulated mostly in genes with unknown functions (44%) but also in genes related to filamentous growth (9%), cell adhesion (7%), metabolic processes (9%) and transcription regulation (6%) (Figure 4-9D). This may indicate that the increased degree of mistranslation induces a strong pressure on the fungal population towards genotype changes at polymorphic locations in genes conferring survivability in the altered proteome landscape where most proteins have serine replaced by leucine.

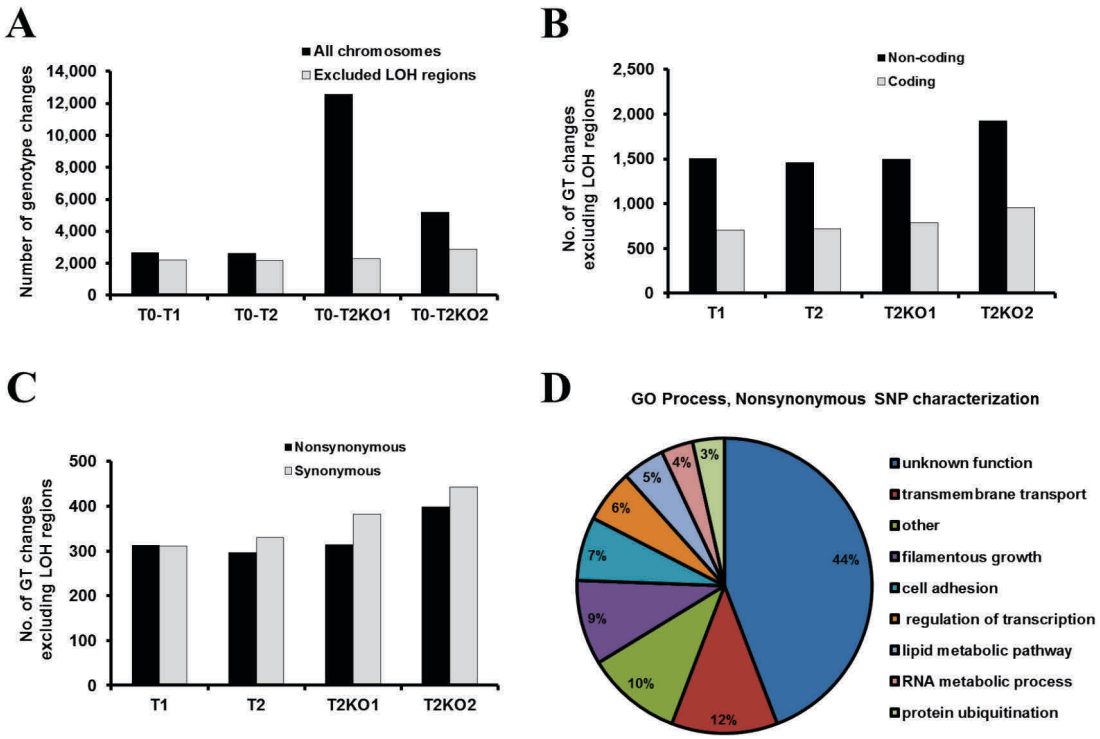


Figure 4-9. Genotype changes from control T0 excluding LOH regions. A. Increasing levels of leucine misincorporation produced strains whose genome contained increasing number of genotype changes at polymorphic locations, compared to T0 (grey bars). This analysis was heavily affected by including the LOH regions (black bars). **B.** In all genomes, most genotype changes outside the LOH regions are located at non-coding positions. **C.** Although rates of nonsynonymous mutations are similar among the different mistranslating genomes, the total number of synonymous genotype changes is slightly higher than nonsynonymous genotype changes. **D.** Gene Ontology (GO) process terms for the unique genes containing nonsynonymous genotype changes identified in T2KO2 compared to control T0. The graph represents the percentage of genes from the total of 86 containing nonsynonymous mutations. GO process characterization was performed using the *Candida* Genome Database (CGD).

Although the majority of the 86 genes with nonsynonymous mutations in T2KO2 possess only 1 genotype change, some genes accumulated more mutations than others. For example, filamentous growth regulators such as *FGR23*, *FGR6-4* and *FGR6-1* and the multidrug resistance factor encoded by *SGE1* had more than 3 nonsynonymous mutations (Table 4-4). Furthermore, several of those genes have an unusual high number of CUG codons, namely *FGR23* with 12 CUGs. To test the hypothesis that genes with higher number of CUGs accumulated higher level of unique SNPs in strain T2KO2, we calculated the correlation coefficient between the number of CUGs and the number of genotype changes using the complete set of 86 genes. Pearson's correlation coefficient was 0.54 or 54%, which means that the two variables have a positive correlation (Figure 4-10A). Since it is more likely to find SNPs in genes that have longer ORF lengths and these usually have more CUG codons, we also tested the correlation between the ORF length and the number of nonsynonymous genotype changes using the complete set of 86 genes. Conversely to the prior comparison, the correlation coefficient was 0.12 or 12%, indicating that the two variables have only a small positive correlation (Figure 4-10A). Therefore, it is more likely that the number of CUGs within an ORF influences the number of nonsynonymous mutations than the ORF length *per se*. Moreover, in strain T2KO2, 81.4% of the genes with nonsynonymous genotype changes have CUGs, while the sequenced genome of strain SC5314 only contains CUG codons in 66% of its genes (Figure 4-10B). Finally, the observation that genes with higher number of CUGs accumulated higher levels of nonsynonymous SNPs is consistent with the higher rate of genotype changes in this most strongly misincorporating strain.

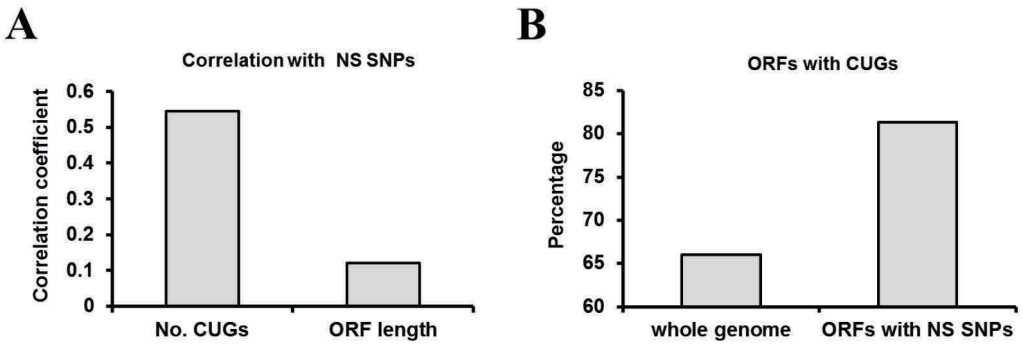


Figure 4-10. T2KO2 nonsynonymous genotype changes from control T0. A. Distribution of CUG codons in T2KO2 genes with nonsynonymous mutations showing that genes with higher number of CUGs accumulated more unique SNPs in strain T2KO2. This correlation was stronger than the relation between ORF length and nonsynonymous genotype changes. Pearson's correlation was used as a measure of relationship between the two variables. **B.** In strain T2KO2, 81.4% of the genes with nonsynonymous genotype changes have CUGs.

Table 4-4. List of T2KO2 genes with higher number of nonsynonymous genotype changes.

ORF ID	No. of GT changes	No. of CUGs	ORF length (nt)	ORF name	Description
orf19.1616	5	12	3,345	FGR23	Filamentous growth regulator
orf19.5543	5	7	2,436	0	-
orf19.1942	5	5	1,884	SGE1	Multidrug resistance factor
orf19.5775	5	4	1,539	0	-
orf19.3490	4	3	1,539	FGR6-4	Filamentous growth regulator
orf19.6027	4	1	1,101	0	-
orf19.48	3	6	4,182	0	-
orf19.5191	3	3	1,539	FGR6-1	Filamentous growth regulator
orf19.1935	3	0	447	0	-

In the five sequenced genomes, we did not observe any shift in codon usage between the strains (Figure 4-11), which indicates that the codon usage is unaffected by degree of mistranslation. Additionally, the frequency of each codon was very similar to the previously published data for SC5314. For example, we detected between 12,514 and 12,576 CUG codons (Figure 4-12A). These numbers are comparable to the 13,074 CUG codons in the genome of the reference SC5314 strain (Butler *et al.*, 2009). When regions of LOH were excluded from the analysis, those numbers of CUG codons lowered to 10,965 and 11,011 (Figure 4-12A).

When codons were analysed in terms of amino acid changes in comparison to the T0 strain, it was evident that strains with high levels of leucine misincorporation had a higher percentage of amino acid changes involving serine and leucine. In strains T1 and T2 only 4.8% and 2.7% of the amino acid changes involved serine and leucine, while in strains T2KO1 and T2KO2 those changes represented 36.0% and 31.3%, respectively (Figure 4-12B). Among the codon changes involving serine and leucine in T2KO1 and T2KO2, the majority were changes from serine to any other amino acid (35 changes in T2KO1 and 42 in T2KO2) or changes from any amino acid to serine (40 and 39 changes, respectively). Changes from leucine to any amino acid were 11 and 16 in T2KO1 and T2KO2, while changes from any amino acid to leucine were 14 in both cases. A smaller fraction of the changes involved a leucine for serine replacement (3 changes in T2KO1 and 9 in T2KO2) or a serine for leucine replacement (10 and 5 changes, respectively) (Table 4-5). Although there was a noteworthy percentage of changes involving serine (therefore the CUG codon) in T2KO1 and T2KO2, the bulk of nonsynonymous SNPs corresponded to changes involving other amino acids (Figure 4-12B). Our data does not explain the high frequency of SNPs at non-CUG sites (non-serine changes), but amino acid replacement at these sites may suppress or minimize protein structural instability generated by leucine misincorporation at CUG sites.

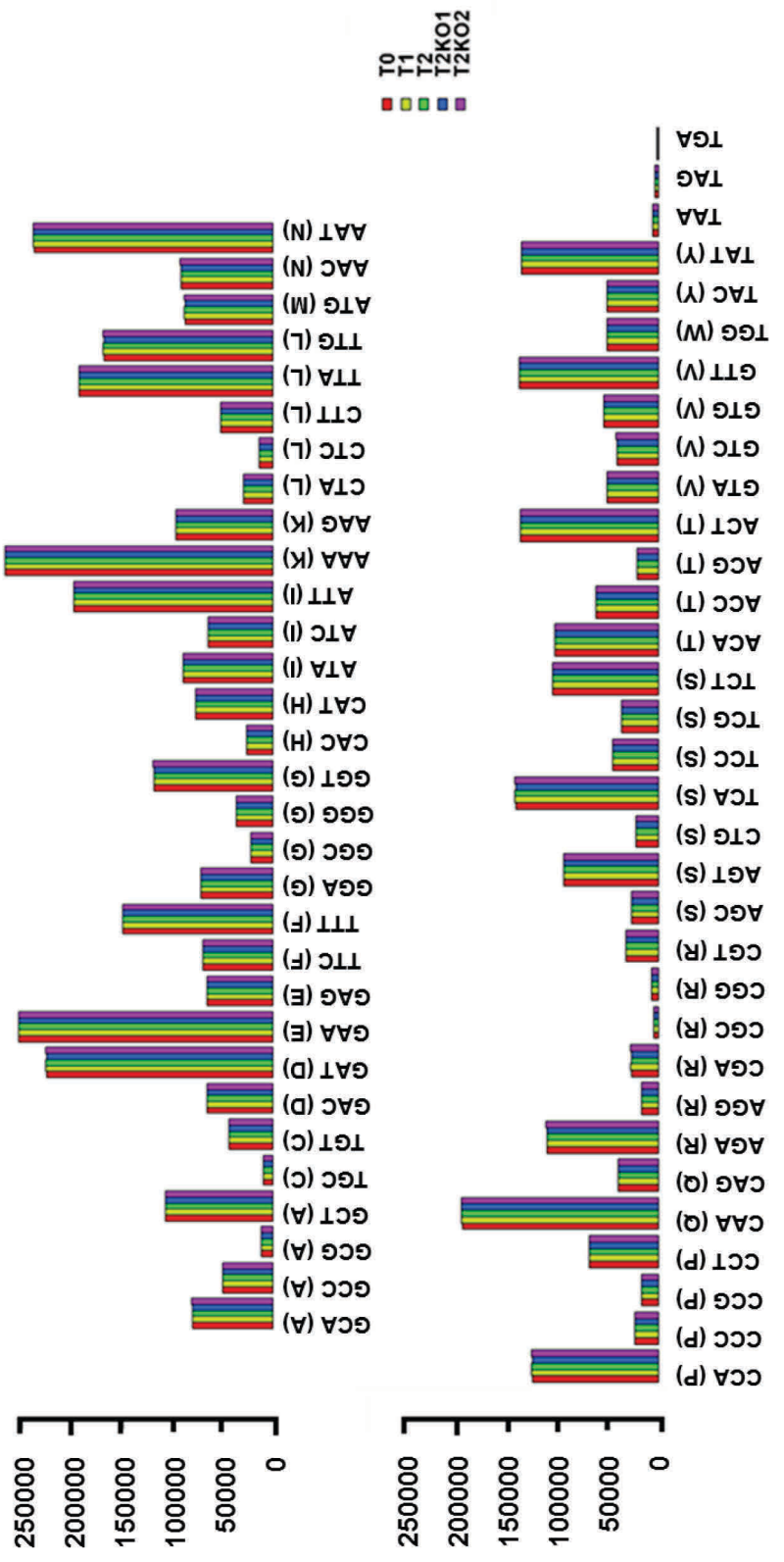


Figure 4-11. The codon usage is unaffected by degree of mistranslation. Each group of bars represents a codon, and the coloured subbars indicate each of the five samples.

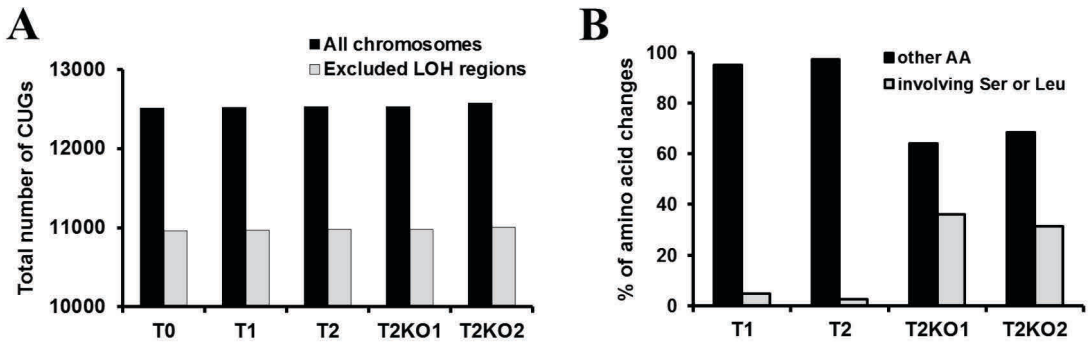


Figure 4-12. Codons detected in the five sequenced genomes. A. Number of CUG codons detected in the complete sequenced genomes (black bars) and excluding the 825,000 ~ 1,136,000 region of chromosome 5 and the 291,000 ~ 1,814,000 region of chromosome R. **B.** Types of amino acid changes between the control strain T0 and mistranslating strains. Codon changes involving serine or leucine are in grey and changes involving any other amino acid are in black.

Table 4-5. Number of amino acid changes involving serine and leucine between control strain T0 and mistranslating strains.

AA changes		Number of codon changes excluding two regions [1)]			
T0	**T1, T2, T2KO1, T2KO2**	**T1**	**T2**	**T2KO1**	**T2KO2**
Leu	**not Leu**	1	1	11	16
not Leu	**Leu**	0	0	14	14
Ser	**not Ser**	7	4	35	42
not Ser	**Ser**	6	2	40	39
Leu	**Ser**	0	0	3	9
Ser	**Leu**	1	1	10	5
Total		15	8	113	125

1) Chr5: 825000 ~ 1136000
ChrR: 291000 ~ 1814000

4.5 Discussion

Genetic diversity is frequently generated during adaptation to stress, and in eukaryotes part of the diversity is produced through recombination and reassortment of alleles during meiosis. Genome rearrangements among fungal species include large deletions between repeated sequences, chromosome breakage and healing and translocations, as well as the gain or loss of chromosomes (Zolan, 1995). Such events have been best characterized in *S. cerevisiae* because of the availability of the complete genome sequence since 1996 which allowed the development of proper molecular tools to study chromosomal changes. For example, during the production of the complete set of gene deletion strains in diploid parental strains, 8% of the strains tested underwent an aneuploidy event and most of them arose from recombination between repeated sequences such as transposons (Hughes *et al.*, 2000b;Dunham *et al.*, 2002). Other fungi also exhibit high levels of chromosome instability, during growth *in vitro* or when propagated in the host. For example, heat shock treatment of *Ustilago hordei* resulted in the deletion of 50 kb near the terminus of a 940 kb chromosome. In *Neurospora crassa*, isolates escaping from het-c heterokaryon incompatibility underwent large deletions of up to 31 kb, including up to 11 ORFs (Xiang and Glass, 2004).

In *C. albicans*, different types of genomic diversity events have been extensively studied and there are evidences that this pathogen may have evolved mechanisms to generate and tolerate extreme genetic changes as a means of adaptation (Selmecki *et al.*, 2010;Forche *et al.*, 2011). Although *C. albicans* has a parasexual life cycle, it reproduces mainly by mitosis and it is unclear how it generates genetic diversity beyond typical genome replication error (Scherer and Magee, 1990;Magee and Magee, 2000;Bennett and Johnson, 2003;Bennett and Johnson, 2005;Butler *et al.*, 2009). In our study, cataloguing the genetic diversity of *C. albicans* strains with different levels of leucine misincorporation at CUG positions provided insights into a unique mechanism of generating polymorphic variation in an essentially clonal organism. The generation of such data required approaches to assay entire genomes at single-base resolution to expose point

mutations and at megabase scale to define structural rearrangements. This last approach included classical tools such as flow cytometry and electrophoretic karyotyping. Results showed that each mistranslating strain was diploid (2N) or close to diploid, with the exception of strains with high level of ambiguity (T1KO1 and T2KO1) that showed a trend to be less than diploid (although there were not evidences of pure haploid isolates). The same strains showed slight changes in the sizes of the smaller chromosomes 5, 6 and 7. Such size changes in small chromosomes have been reported before and they have been linked to the presence of a major repeat sequence (MRS), a tract of repeats that varies in size from 10 kilobase pairs to 100 kilobase pairs (Magee, 2007). Chromosomes 5 and 6 have one copy of the major repeat sequence (MRS) while chromosome 7 has two copies of the MRS. Because MRSs are the largest nontelomeric, homologous sequences present in each chromosome, they are usually related with highly frequent translocations involving these chromosomes (Iwaguchi *et al.*, 2001;Lephart and Magee, 2006). Also, expansion and contraction of MRSs in small chromosomes contribute to chromosome length polymorphisms along with chromosome truncations and deletions of chromosome segments (Lephart *et al.*, 2005). Thus, ambiguous CUG decoding induced genome destabilization, which may have occurred through several processes, namely translocation, deletion, amplification and chromosome loss or gain.

Full genome sequencing using Illumina technology produced high-depth genome sequences for strains T1, T2, T2KO1 and T2KO2 and their parental strain T0. The genome sequence of the T0 strain was essential to pinpoint mutations because it is the naturally ambiguous strain that served as template against which the genomes of the descendent strains were compared. Whole-genome sequencing confirmed the tRNA gene integrations and knockouts but we were unable to identify indels, copy number variants or large chromosome alterations. However, one cannot exclude them since they are difficult to score using the short Illumina sequencing reads. This difficulty may be surpassed in future analyses of the short read sequencing data because new approaches for detection of copy number polymorphisms and determination of rearrangement structures in yeast are currently under development (Ruffalo *et al.*, 2012;Teo *et al.*, 2012;Neuman *et*

al., 2012;Henson *et al.*, 2012). Recently, Araya and colleagues (2010) compiled the read-depth data from evolved and parental genome sequencing data sets of yeast strains and calculated the ratio in read-depth for each base across the genome. These data was then used to perform circular binary segmentation (CBS) with *DNAcopy* (Venkatraman and Olshen, 2007) that formulates the problem of copy-number variation (CNV) analysis as a change-point detection problem. The algorithm searches for change-points by partitioning the genome and performing a maximal *t*-test between segment ratios, joining segments with comparable ratios. This segmentation approach detected a 5x amplification of a segment of chromosome 2 in the evolved yeast strain which corresponded to *SUL1*. The result was confirmed using tiling array data which proved that read depth-based copy number analysis is reliable and can be used in copy number polymorphisms estimations (Araya *et al.*, 2010).

Another aspect of the short read sequencing data that can be improved in future analyses concerns the unmapped sequences of the mistranslating strains. In our data, the majority of the reads had perfect sequence identity with the reference regions, whereas a small portion (~3.3%) had frequent differences, indicating a different sequence source not present in the reference sequence of SC5314. It may be interesting to see what those unmapped reads are since they may fall at breakpoints of structural variation. For instance, in the work mentioned above, unmapped reads were assembled into contigs using *Velvet*, a short-read *de novo* assembly algorithm (Zerbino and Birney, 2008). These contigs were then BLAT-aligned to the reference genome sequence in an attempt to detect signatures of novel structures in the genome (Araya *et al.*, 2010).

Although short sequencing reads represent a challenge in the assembly of highly rearranged genome segments or in regions of structural variation, the high fold coverage that they provide is highly useful for single nucleotide polymorphisms (SNPs) detection. *C. albicans* has one of the most polymorphic genomes of the thousands of genomes sequenced to date with more than 48,000 SNPs in strain SC5314, giving 1 SNP per 390 bases (Jones *et al.*, 2004;Braun *et al.*, 2005;Butler *et al.*, 2009), and this is even higher in strain WO-1. In this case, between 87,000 and 89,000 SNPs were identified, suggesting that only a very

small number of SNPs are actually shared between the two genomes (Butler *et al.*, 2009). In our study, we detected between 45,297 and 55,429 SNPs in the five sequenced strains which match the sequencing based predictions for SC5314, which was expectable because all mistranslating strains were derived from the SC5314. When SNPs were compared to the naturally ambiguous T0 strain, the rate of variation increased with increasing degree of mistranslation. In fact, the reverted strain T2KO2 had 600 more SNPs than the other strains. The increased rate of changes from control is accumulated mostly in the non-coding region of the DNA. In *C. albicans*, 63.0% of the genome constitutes coding regions and the lower proportion of SNPs in these regions expected by a uniform random rate of SNP occurrences can be explained by the increased risk for a functionally disruptive mutation in those regions. A mutation in a non-coding region is much less likely to affect the organism's phenotype than one in a coding region. Nevertheless, mutations in non-coding regions do have a cost because they play a vital role in the regulation of gene expression. Some of the non-coding DNA provides proper framing for translation of proteins (Trifonov, 1989). In addition, there are 5’ and 3’ untranslated regions (UTRs) consisting of promoters and terminators that regulate gene translation since certain trans-acting binding proteins bind to the 3' and 5' UTRs of proximal and distal genes to regulate their translation (Mazumder *et al*., 2003;Lackner and Bahler, 2008;Araujo *et al*., 2012). Another role for the 3' and 5' UTRs is to regulate the rate of mRNA decay (Raue, 1994;Newbury, 2006), and mutations in these non-coding regions may affect gene expression through mRNA stability.

Our data also showed that genes with higher number of CUGs accumulated higher level of SNPs in strain T2KO2. One possible explanation is that codon ambiguity increases the rate of genotype changes at polymorphic locations to suppress or minimize protein structural instability generated by amino acid misincorporations. The selection of alternative amino acids at non-ambiguous sites may re-tune protein-protein, protein-RNA and protein-DNA interactions and networks destabilized by misincorporations. Although the reverted strain T2KO2 presented more SNPs than any other strain, no alteration in codon usage was noticeable. Moreover, the frequency of each codon was very similar to the

previously published data for SC5314 whose genome encodes 13,074 CUG codons that are spread over 66% of its 6,438 genes (Gomes *et al.*, 2007;Butler *et al.*, 2009). Previous studies have shown that CUG reassignment is 270 ± 25 million years (My) old, evolved through CUG ambiguity and that the latter erased the CUG codons from the genome of the CTG clade ancestor (Massey *et al.*, 2003;Santos *et al.*, 2004;Butler *et al.*, 2009). Those studies also showed that most CUGs present in extant CTG clade species re-emerged gradually over the last 270 My through mutations of codons assigned to serine or amino acids with similar chemical properties (Massey *et al.*, 2003;Silva *et al.*, 2004). X-ray crystallography and molecular modeling of *C. albicans* proteins demonstrated that these new CUGs are located in protein sites where serine and leucine incorporations are tolerated by protein structure (Rocha *et al.*, 2011), despite their distinct chemical properties. The data presented here validates those hypothesis and further show that such mutational events can occur in a short time span.

Whole genome sequencing also allowed for the detection of loss of heterozygosity (LOH) events in the genomes of highly mistranslating strains, namely T2KO1 and T2KO2. These results suggested that these strains may be under a higher level of stress because Forche (2011) showed that LOH rates increased with increased levels of stress and that different stress conditions caused different types of LOH events (Forche *et al.*, 2011). In fact, increasing rates of LOH in the presence of stress likely facilitates the adaptation of *C. albicans* to stressful environments since they can expose hidden phenotypes associated with recessive alleles. Although heterozygosity prevails at most loci, some strains are known to carry homozygous regions. For example, in strain SC5314 both ends of chromosome R, most of the right arm of chromosome 3, and the left arm of chromosome 7 are nearly homozygous. In strain WO-1 both arms of chromosome 5 and most of the left arm of chromosome 2, the left arm of chromosome 6, the right arm of chromosome 3, and the right arm of chromosome 1 are also nearly homozygous (Butler *et al.*, 2009;Selmecki *et al.*, 2010). In this study, T2KO1 isolate had complete loss of heterozygosity for the entire chromosome R and near-complete loss of heterozygosity (LOH) in a 300 kb region of chromosome 5 (825,000 - 1,136,000). The same region of chromosome 5 was

also affected by LOH in the T2KO2 strain. This particular region on chromosome 5 contains genes involved in filamentous growth, cell adhesion and stress response, suggesting that CUG ambiguity does provide an unanticipated mechanism to generate genetic diversity in genes associated with *C. albicans* pathogenesis and drug resistance. Although, the location of the LOH events in our study is different from the previous reports, it is interesting to notice that some of the genes in the LOH region are crucial in the calcineurin transduction pathway. For example, gene *CNB1* codifies the regulatory subunit of calcineurin and *MID1* encodes a calcium channel (Sanglard *et al*., 2003a). The exact effect of the LOH in *CNB1* and *MID1* is not known but protein phosphatase calcineurin is part of a well described mechanism of drug resistance (Cruz *et al.*, 2002;Sanglard *et al.*, 2003a). In normal cells, calcineurin is in an inactive state and external signals are required to activate calcineurin and its phosphatase activity. One of the known activators of this pathway is membrane stress that allows the release of Ca^{2+} stores in the cell through the calcium channel composed of Mid1 and Cch1. Ca^{2+} then binds calmodulin, which activates calcineurin. One of the main targets of this phosphatase is the transcription factor Crz1 that once it is dephosphorylated enters the nucleus and induces expression of a number of genes, many of which encode proteins with cell wall related functions (Cruz *et al.*, 2002).

As for LOH on the entire chromosome R in T2KO1, our data suggests that independent isolates of the same strain are different, i.e., not all T2KO1 isolates possess a homozygous chromosome R. This is in line with previous studies from Rustchenko showing that chromosome R is twice as unstable as the other chromosomes (Rustchenko, 2007). The explanation for the dynamic nature of chromosome R could be related to the presence of both homologs of the cluster of tandemly organized rRNA units (Iwaguchi *et al.*, 1992;Rustchenko *et al.*, 1993). Furthermore, a study performed on unstable mutants derived from the *C. albicans* 3153A laboratory strain demonstrated that chromosome R was highly unstable (Wellington *et al.*, 2006). Authors suggested that the instability within those strains could simply reflect an increased level of recombinogenic events in the genome, as multiple tandem direct repeats obviously favour such events. In this respect,

chromosome R may be a sensitive measure of a general instability, which in this case implies that strain T2KO1 is highly unstable.

Other aspects of the genetic diversity detected with the sequencing data were the dissimilar types of LOH in the two chromosomes affected. While chromosome 5 had a segmental LOH, chromosome R presented a whole-chromosome LOH. Several models have been proposed for how aneuploidy arises and how they can result in whole-chromosome and segmental LOH. One of these models involves the formation of polyploidy cells that are usually generated by mating, protoplast fusion, or defects in cell cycle events, such as mitosis or replication. Then, polyploid cells lose extra chromosomes via nondisjunction events in a process designated as concerted chromosome loss (CChrL). It is unclear whether CChrL is regulated, similar to aberrant meiosis, or whether it is a breakdown in centromere/spindle functions and/or in the checkpoints that monitor them. Either way, this process results in whole-chromosome aneuploidy. Other mechanism that has been associated with the generation of aneuploidy and LOH involves mitotic nondisjunction, the failure of sister chromatids to separate during mitosis. This phenomenon yields two aneuploid cells: one that is monosomic and thus necessarily homozygous for that chromosome and the other that is trisomic and heterozygous (with a 2:1 allelic ratio) for the same chromosome. Whole-chromosome homozygosis arises by re-replication of the monosomic chromosome or by loss of the heterozygous homolog in a trisomic strain (Selmecki *et al.*, 2010). However, the mechanisms by which stress conditions allow these events to bypass cell cycle checkpoints have been poorly explored. Forche (2011) showed that different stress conditions yield different levels of LOH and hypothesized that the types of genetic diversity are directly related to the type of DNA damage inflicted by the stress condition (Forche *et al.*, 2011). For example, temperature stress causes chromosome loss and aneuploidy (Hilton *et al.*, 1985;Bouchonville *et al.*, 2009;Forche *et al.*, 2011), and it is speculated that they occur through the limitation in heat shock protein chaperones and co-chaperones, namely Sgt1, Sti1 and Cdc37, which are essential in the assembly of functional kinetochores and spindle pole bodies/centrosomes and in the function of mitotic checkpoints (Lange *et al.*, 2000;Felts *et al.*, 2007). Additionally, fluconazole treatment is known to

cause whole-chromosome LOH (Selmecki *et al.*, 2006;Selmecki *et al.*, 2009) and the prevalent model assumes that changes in ergosterol biosynthesis affect nuclear membrane fluidity, which in turn impact the function of the spindle pole bodies embedded in the nuclear membrane and the ability of cytoplasmic dynein to properly associate with the cortex and execute proper nuclear separation (Finley *et al.*, 2008;Webster *et al.*, 2009). This affects chromosome segregation mechanisms and result in polyploidy and/or aneuploidy (Forche *et al.*, 2011). The mechanism by which increased mistranslation causes whole-chromosome and segmental LOH is still not clear, but impairment of DNA repair pathways may be a possibility. The other is that mistranslated proteins deviate Sgt1, Sti1 and Cdc37 from the kinetochores.

If one takes into consideration the Specific Codon Usage (SCU), which measures the relative frequency of CUGs normalized to serine abundance (Ikemura, 1982;Santos *et al.*, 2011), it is clear that CUG usage is positively biased in genes of the Golgi to endosome transport, DNA repair, regulation of redox homeostasis, chromatin silencing and modification, mRNA processing, cell growth and/or maintenance and budding (Table 4-6). Among such categories, DNA repair genes have been described as being crucial for genome stability maintenance (Legrand *et al.*, 2007;Legrand *et al.*, 2008;Legrand *et al.*, 2011;Hoot *et al.*, 2011). Legrand and colleagues (2007) showed that defects in double-strand break repair (DSBR) and mismatch repair (MMR) increase genome instability, specifically defects in the MMR pathway lead to point mutations and instability of repetitive DNA tracts, while defects in the DSBR pathway cause gross chromosome rearrangements (Drotschmann *et al.*, 2000;Legrand *et al.*, 2007). Since most of the genes involved in these processes have CUG codons, it is likely that increasing leucine misincorporation at those positions interferes with protein function and consequently with the DNA repair pathway. For example, *RAD50* and *MRE11* genes that encode DNA double-strand break repair factors have 10 and 5 CUGs, respectively. *PMS1* and *MSH2* genes that encode DNA mismatch repair factors have 7 and 4 CUGs, respectively. Similarly, disruption of cell cycle checkpoints are known to cause genome rearrangements and genes encoding such proteins also have CUG codons. Two genes of the S-phase checkpoint are

especially important, *MEC1* and *SGS1* (Legrand *et al.*, 2011), and both have 13 CUGs. In our highly ambiguous strains, the function of *MEC1* and *SGS1* may be partially compromised and this can be associated to the LOH events detected in these strains. The work of Legrand (2011) has also shown that strains with deletion of the *MEC1* gene have a high incidence of LOH events, suggesting that these events are detected and repaired by the MEC1 pathway. Therefore, loss of the MEC1-dependent checkpoint increases either the LOH incidence, or decreases the probability of repair (Legrand *et al.*, 2011). A similar effect was previously detected with *MRE11* and *RAD50* mutants (Legrand *et al.*, 2007), in which deletion of either gene led to an increase in overall LOH frequency.

Table 4-6. Functional categories (GO terms) of genes with highest SCU_{CUG}.

Cellular process	No. of genes	SCU_{CUG}
Golgi to endosome transport	11	0.090
DNA repair	42	0.070
Regulation of redox homeostasis	10	0.068
Golgi to vacuole transport	15	0.064
Chromatin silencing	22	0.062
mRNA splicing	54	0.059
Golgi to plasma membrane transport	16	0.057
Mitosis	36	0.057
Cytokinesis	16	0.056
Protein targeting	37	0.056
Chromatin modification	11	0.055
Cyclin catabolism	10	0.054
Budding	19	0.053
G1 phase of mitotic cell cycle	17	0.052
mRNA processing	41	0.052
Signal transduction	11	0.051
Cell growth and/or maintenance	16	0.051
Autophagy	18	0.050
Protein amino acid phosphorylation	23	0.050
Regulation of transcription	32	0.050

Candida albicans genes were grouped by their Gene Ontology (GO) terms and the average CUG usage (SCU_{CUG}) of each category was calculated. CUGs are enriched in specific GO terms.
The impact of CUG mistranslation is maximal in these genes (adapted from (Santos *et al.*, 2011).

4.6 Conclusions

Several studies have highlighted the high plasticity of the *C. albicans* genome as a contributor to drug resistance through loss of heterozygozity (LOH) in *loci* associated to drug resistance. Other studies showed that chromosomal rearrangements amplify gene copy number of drug resistance genes. How *C. albicans* tolerates such genome changes and whether they represent a global adaptive response of the pathogen to stress remains to be fully clarified. Our data show that codon ambiguity increases the rate of genotype changes at polymorphic locations and generates loss of heterozygosity in a region of chromosome 5 that contains genes related to drug resistance. Our discovery that CUG ambiguity has direct impact on genetic diversity adds therefore a new dimension to the study of genome evolution, phenotypic variation, ecological adaptation and drug resistance and provides strong evidence for subtle and yet remarkable power of codon ambiguity in shaping gene evolution.

5. General discussion

5.1 General discussion

The genetic code defines highly conserved codon-to-amino acid assignment rules that govern the translation of genes into proteins across the 3 domains of life. Therefore, understanding the evolution of the genetic code is fundamental to elucidate the transition of life from the RNA world to the DNA-protein world. The near invariability of these rules generalized the idea that the code is frozen and immutable (Crick, 1968), however various fungi, ciliates and bacteria have reassigned various codons, showing that codons have some degree of assignment flexibility (Knight *et al.*, 2001). Both neutral and non-neutral theories have been proposed to explain codon reassignments (Osawa and Jukes, 1989;Schultz and Yarus, 1994), however experimental data to support or refute them is scarce and genetic code alterations remain an intriguing biological puzzle. The “codon capture theory” postulates that under strong CG- or AT- pressure, certain codons can disappear, which allows the loss of the tRNAs that decode them. At a later stage, erased codons can reappear and become reassigned by mutant tRNAs from different isoacceptor families (Osawa and Jukes, 1989). This theory is supported by the presence of unassigned codons in *Mycoplasma capricolum* (CGG) and *Micrococcus luteus* (AGA and AUA), which have 75% of AT and GC, respectively (Osawa *et al.*, 1992). Since codons are erased from the genome, this theory is neutral as it does not introduce amino acid alterations into proteins upon codon reappearance in genes. On the other hand, the “ambiguous intermediate theory” postulates that structural changes in translational factors are the key element in any genetic code alteration (Schultz and Yarus, 1994). Such structural changes can occur in tRNAs, allowing them to recognize near-cognate codons, producing codon ambiguity. Since codon ambiguity destabilizes the proteome, it is likely toxic and can only be selected if it brings selective advantages to the organism, but the theory does not explain what kind of advantages may result from ambiguous codon decoding. The reassignment of the CUG codon from leucine to serine in fungi of the *Candida*, *Debaryomyces* and *Lodderomyces* genera (CTG clade) (Ohama *et al.*, 1993;Santos and Tuite, 1995;Sugita and Nakase, 1999) supports this theory. Extant *C. albicans* strains

incorporate 97% of serine and 3% of leucine into proteins under standard laboratory conditions (Gomes *et al.*, 2007).

The study of the evolution of the genetic code is important to many fields of basic and applied research, from evolutionary and genome biology to synthetic biology and biotechnology. In fact, infiltration of natural and synthetic amino acids at nonsense codons and the engineering of *E. coli* strains and mammalian cells that tolerate high levels of amino acids misincorporation have been achieved and proteins with novel functionalities have been produced (Bacher and Ellington, 2001;Wang *et al.*, 2009;Mukai *et al.*, 2010;Young and Schultz, 2010;Hoesl and Budisa, 2012). Despite this, the failure to completely alter codon identity generalized the idea that the genetic code is largely non-tractable experimentally. In this thesis, we took advantage of the atypical genetic code alteration of the main human pathogen *Candida albicans* (Santos and Tuite, 1995;Gomes *et al.*, 2007) to push CUG ambiguity to its limit and to demonstrate that codon ambiguity provides an effective mechanism to alter the genetic code. To achieve this, we have applied a combination of genetic engineering and molecular biology to alter the identity of the CUG codon and we used genomics, phenomics and transcriptomics to characterize a series of novel fungal strains with increasing levels of CUG ambiguity.

To increase CUG ambiguity we have inserted one (strain T1) or two (strain T2) copies of a yeast leucine $tRNA_{CAG}^{Leu}$ gene into the genome of *C. albicans*. We also knocked out one or both copies of the chromosomal *C. albicans* serine $tRNA_{CAG}^{Ser}$ gene in strains T1 and T2, producing strains T1KO1 and T2KO1 or T2KO2. This allowed us to produce strains of *C. albicans* that incorporate 20.61%, 50.04%, 67.29%, 80.84% and 98.46% of leucine at an atypical serine CUG codon on a proteome scale. This tolerance to high levels of leucine misincorporation is a direct consequence of the location of CUG codons at sites that can accommodate both leucine and serine (Rocha *et al.*, 2011). Although CUG codons occur in 66% of *C. albicans* protein-coding genes, 90% of these CUG codons are located in positions where either aliphatic leucine or polar serine can be inserted without causing disruption of protein structure and function (Rocha *et al.*, 2011). In the

remaining 10% of CUG codons, the replacement of polar serines by hydrophobic leucines in proteins should have consequences for their tridimensional structure and function. Interestingly, these CUGs are located in genes involved in virulence and pathogenesis, namely, biofilm formation, morphogenesis, and mating (Rocha *et al.*, 2011), suggesting that CUG ambiguity may have functional roles in *C. albicans* virulence.

Our strains also represent a unique resource for geneticists, microbiologists and evolutionists interested in the study of the role of proteome stability and codon ambiguity in genome evolution and cell biology and it will be important to explore the biology of this proteome instability in future studies. For example, it would be interesting to have a comprehensive picture of variation in protein concentration in response to different levels of CUG mistranslation and compare such data with mRNA levels. This type of analysis has been carried out previously by Vogel and colleagues (Vogel and Marcotte, 2008;Vogel *et al.*, 2010;Vogel *et al.*, 2011;Vogel and Marcotte, 2012) using yeast subjected to oxidative stress. Using an Orbitrap mass detector (shotgun proteomics, in which cellular proteins are enzymatically digested and the resulting peptides are analysed by nanoflow chromatography and high resolution tandem mass spectrometry), they provided the first time-resolved concentration measurement of 1,900 yeast proteins during the 2 h after diamide-induced oxidative stress. Integration of proteomic measurements with transcription, translation and other regulatory data allowed the characterization of the general and specific proteome response to oxidative stress, highlighting possible regulatory mechanisms and their targets (Vogel *et al.*, 2011). Also, Schwanhausser and co-workers (2011) estimated protein degradation rates by isotopically labeling newly synthesized proteins that were subjected to shotgun proteomics (Schwanhausser *et al.*, 2011). This approach could be applied to our mistranslating strains in order to determine protein turn-over rates, particularly in strain T2KO2 that completely reassigned the CUG codon from serine to leucine. This should allow one to identify *C. albicans* proteins whose turnover and function is regulated by CUG ambiguity.

We have also demonstrated that codon ambiguity is a potent generator of phenotypic diversity of high adaptation potential. We have phenotyped the ambiguous strains in presence of various nutritional substrates, environmental stressors and antifungals and we showed that increasing levels of CUG ambiguity induced production of arrays of novel colony morphotypes, high cell population heterogeneity, high tolerance to common antifungals and resistance to various environmental stressors. This data is in agreement with previous work from our laboratory which showed that high CUG ambiguity produces phenotypic diversity (Santos *et al.*, 1999;Miranda *et al.*, 2007) and in the case of ambiguous *S. cerevisiae* selective advantages under environmental stress conditions (Santos *et al.*, 1996;Santos *et al.*, 1999;Silva *et al.*, 2007). A number of other laboratories also engineered codon ambiguity in *E. coli* and confirmed that it results in important selective advantages that allow cells to explore new environmental niches or overcome severe stress situations (Bacher and Ellington, 2001;Pezo *et al.*, 2004;Bacher *et al.*, 2005;Nangle *et al.*, 2006;Bacher *et al.*, 2007). Hence, it would be interesting to identify environmental conditions where mistranslation is favorable and use them in forced evolution experiments to gradually increase codon ambiguity and ultimately achieve codon reassignment. This would provide a powerful strategy for creating organisms with alternative genetic codes that could produce new proteins containing novel amino acid chemistries. Nevertheless, how codon ambiguity produces phenotypic variability is still unclear. Likely possibilities are accumulation of DNA mutations, metabolic changes and deregulation of gene expression or stress signaling pathways.

In order to better understand how codon ambiguity produces phenotypic diversity, we have profiled the transcriptome of the engineered strains using DNA microarrays. Surprisingly, gene expression deregulation was small and the typical environmental stress response (ESR) was not induced (Gasch *et al.*, 2000;Causton *et al.*, 2001), highlighting the remarkable adaptation and tolerance of *C. albicans* to CUG ambiguity. Moreover, only the high misincorporating strains T2 and T2KO1 exhibited a typical transcriptional response activated by the Hog1 SAPK pathway, which is the small core stress response of *C. albicans* (Smith *et*

al., 2004;Enjalbert *et al.*, 2006). The observation that this mild response was only activated at very high levels of ambiguity (more than 67%) is in line with the fact that *C. albicans* tolerates well leucine misincorporation at CUGs. The surprising lack of activation of the stress response in *C. albicans* was in sharp contrast with its activation in yeast where partial reconstruction of the genetic code alteration unveiled major gene expression deregulation (Santos *et al.*, 1999;Silva *et al.*, 2007). This result provides fascinating evidence for major differences between the stress responses of both microorganisms and raises the hypothesis that CUG reassignment reprogrammed the *C. albicans* stress response by altering the gene networks that regulate it. This hypothesis is supported by functional divergence of the Msn2/4p master regulators of the stress response between *S. cerevisiae* and *C. albicans* (Nicholls *et al.*, 2004). It would be interesting in future studies to fully characterize the transcriptional network controlling the stress response in highly mistranslating *C. albicans* strains, using whole-genome transcriptional and Chip-Chip location profiling (Hogues *et al.*, 2008). The appearance or disappearance of transcription factor binding motifs in the promoter regions of genes would be highly informative about transcriptional divergence between yeast and *C. albicans* (Ihmels *et al.*, 2004;Hogues *et al.*, 2008).

We have also analysed the contribution of DNA mutations to the phenotypic diversity observed in our strains. Previous studies have shown that a single mutation in a $tRNA^{Gly}_{GCC/U}$ gene that produced a tRNA that misincorporated glycine at aspartate codons (GAU/C) (Slupska *et al*., 1998;Dorazi *et al*., 2002) and that mutations in tRNA modifying enzymes produced a hypermutagenesis phenotype (Slupska *et al.*, 1996;Murphy and Humayun, 1997;Zhao *et al.*, 2001). Re-sequencing of the genome of the *C. albicans* misincorporating strains at very high genome coverage (>99%) confirmed that codon ambiguity is also mutagenic in eukaryotes. Our results showed a clear accumulation of increasing levels of SNPs and loss of heterozygozity (LOH) in the higher misincorporating strains. This data unveiled a link between codon ambiguity and genome mutational load, although the mechanism by which increased mistranslation causes mutations is still not clear. Disruption of DNA repair pathways may be a possibility because

genes within this category contain CUGs. Therefore, it would be important to characterize the activity of DNA repair enzymes in the engineered strains using phenotypic assays, namely sensitivity to oxidizing agents, alkylating agents and ultraviolet (UV) light (Legrand *et al.*, 2007;Legrand *et al.*, 2011). It would also be interesting to delete both copies of *RAD50* and *MRE11* (DNA double-strand break repair factors) in the misincorporating strains and monitor genome stability either by sequencing or aCGH to check if our strains acquire a hypermutagenesis phenotype.

This thesis also demonstrated that genes containing CUG codons accumulated more SNPs than those lacking CUGs and that SNPs accumulated in genes involved in filamentous growth, cell adhesion, transport and cell wall biosynthesis. This suggests that CUG ambiguity provides an unantecipated mechanism to generate genetic diversity in genes associated with *C. albicans* pathogenesis. We have demonstrated previously that the original CUG codons were erased from the genome of the *C. albicans* ancestor and that extant CUGs appeared recently and encode protein residues in sites that tolerate both serine and leucine (Massey *et al.*, 2003;Rocha *et al.*, 2011). Our new data strongly suggests that *C. albicans* CUGs have functional roles beyond the classical codon roles in gene translation.

5.2 Conclusions

We have demonstrated for the first time that ambiguous codons are reassignable, validating old hypotheses proposed by Carl Woese in 1965 and by Michael Yarus in 1992. Carl Woese postulated in his “Translation Error Hypothesis” that ambiguous codons were common in the primordial translation system, that the proteins produced were “statistical proteins” and that ambiguous codons were easily reassignable (Woese, 1965). Yarus and Shultz proposed that such codon ambiguity could also explain the evolution of natural codon reassignments in extant organisms with highly complex genomes (Schultz and Yarus, 1994). Our discovery that codon ambiguity generates genetic and phenotypic diversity of high adaptation potential shows that the proteome chaos generated by codon ambiguity does not impede selection of codon reassignment. The effect of codon ambiguity on phenotypic and genetic diversity adds therefore a new dimension to the study of genome evolution, phenotypic variation, ecological adaptation, drug resistance and human disease and provides strong evidence for subtle and yet remarkable power of codon ambiguity in shaping gene evolution.
It is generally assumed that mutations in DNA alter protein structure. We have demonstrated that mutations in proteins alter DNA primary structure, a clear violation of the central dogma of biological information flow from DNA to proteins. The implications of this discovery are wide and deep, the future will tell its biological relevance.

6. References

Ahmad,A., Kabir,M.A., Kravets,A., Andaluz,E., Larriba,G., and Rustchenko,E. (2008). Chromosome instability and unusual features of some widely used strains of Candida albicans. Yeast *25*, 433-448.

Allmang,C. and Krol,A. (2006). Selenoprotein synthesis: UGA does not end the story. Biochimie *88*, 1561-1571.

Allmang,C., Wurth,L., and Krol,A. (2009). The selenium to selenoprotein pathway in eukaryotes: more molecular partners than anticipated. Biochim. Biophys. Acta *1790*, 1415-1423.

Ambrogelly,A., Palioura,S., and Soll,D. (2007). Natural expansion of the genetic code. Nat. Chem. Biol. *3*, 29-35.

Antonellis,A. and Green,E.D. (2008). The role of aminoacyl-tRNA synthetases in genetic diseases. Annu. Rev. Genomics Hum. Genet. *9*, 87-107.

Araujo,P.R., Yoon,K., Ko,D., Smith,A.D., Qiao,M., Suresh,U., Burns,S.C., and Penalva,L.O. (2012). Before It Gets Started: Regulating Translation at the 5' UTR. Comp Funct. Genomics *2012*, 475731.

Araya,C.L., Payen,C., Dunham,M.J., and Fields,S. (2010). Whole-genome sequencing of a laboratory-evolved yeast strain. BMC. Genomics *11*, 88.

Arbour,M., Epp,E., Hogues,H., Sellam,A., Lacroix,C., Rauceo,J., Mitchell,A., Whiteway,M., and Nantel,A. (2009). Widespread occurrence of chromosomal aneuploidy following the routine production of Candida albicans mutants. FEMS Yeast Res. *9*, 1070-1077.

Ardell,D.H. and Sella,G. (2001). On the evolution of redundancy in genetic codes. J. Mol. Evol. *53*, 269-281.

Ardell,D.H. and Sella,G. (2002). No accident: genetic codes freeze in error-correcting patterns of the standard genetic code. Philos. Trans. R. Soc. Lond B Biol. Sci. *357*, 1625-1642.

Arnez,J.G. and Moras,D. (1997). Structural and functional considerations of the aminoacylation reaction. Trends Biochem. Sci. *22*, 211-216.

Auesukaree,C., Damnernsawad,A., Kruatrachue,M., Pokethitiyook,P., Boonchird,C., Kaneko,Y., and Harashima,S. (2009). Genome-wide identification of genes involved in tolerance to various environmental stresses in Saccharomyces cerevisiae. J. Appl. Genet. *50*, 301-310.

Augustin,M.A., Reichert,A.S., Betat,H., Huber,R., Morl,M., and Steegborn,C. (2003). Crystal structure of the human CCA-adding enzyme: insights into template-independent polymerization. J. Mol. Biol. *328*, 985-994.

Avery,S.V., Howlett,N.G., and Radice,S. (1996). Copper toxicity towards Saccharomyces cerevisiae: dependence on plasma membrane fatty acid composition. Appl. Environ. Microbiol. *62*, 3960-3966.

Bacher,J.M., De Crecy-Lagard,V., and Schimmel,P.R. (2005). Inhibited cell growth and protein functional changes from an editing-defective tRNA synthetase. Proc. Natl. Acad. Sci. U. S. A *102*, 1697-1701.

Bacher,J.M. and Ellington,A.D. (2001). Selection and characterization of Escherichia coli variants capable of growth on an otherwise toxic tryptophan analogue. J. Bacteriol. *183*, 5414-5425.

Bacher,J.M., Hughes,R.A., Tze-Fei,W.J., and Ellington,A.D. (2004). Evolving new genetic codes. Trends Ecol. Evol. *19*, 69-75.

Bacher,J.M., Waas,W.F., Metzgar,D., De Crecy-Lagard,V., and Schimmel,P. (2007). Genetic code ambiguity confers a selective advantage on Acinetobacter baylyi. J. Bacteriol. *189*, 6494-6496.

Baranov,P.V., Gesteland,R.F., and Atkins,J.F. (2002). Recoding: translational bifurcations in gene expression. Gene *286*, 187-201.

Barelle,C.J., Manson,C.L., MacCallum,D.M., Odds,F.C., Gow,N.A., and Brown,A.J. (2004). GFP as a quantitative reporter of gene regulation in Candida albicans. Yeast *21*, 333-340.

Barnett,D.W., Garrison,E.K., Quinlan,A.R., Stromberg,M.P., and Marth,G.T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics. *27*, 1691-1692.

Barrell,B.G., Bankier,A.T., and Drouin,J. (1979). A different genetic code in human mitochondria. Nature *282*, 189-194.

Barrick,J.E., Yu,D.S., Yoon,S.H., Jeong,H., Oh,T.K., Schneider,D., Lenski,R.E., and Kim,J.F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature *461*, 1243-1247.

Bennett,R.J. and Johnson,A.D. (2003). Completion of a parasexual cycle in Candida albicans by induced chromosome loss in tetraploid strains. EMBO J. *22*, 2505-2515.

Bennett,R.J. and Johnson,A.D. (2005). Mating in Candida albicans and the search for a sexual cycle. Annu. Rev. Microbiol. *59*, 233-255.

Bensen,E.S., Martin,S.J., Li,M., Berman,J., and Davis,D.A. (2004). Transcriptional profiling in Candida albicans reveals new adaptive responses to extracellular pH and functions for Rim101p. Mol. Microbiol. *54*, 1335-1351.

Berman,J. and Hadany,L. (2012). Does stress induce (para)sex? Implications for Candida albicans evolution. Trends Genet. *28*, 197-203.

Berman,J. and Sudbery,P.E. (2002). Candida Albicans: a molecular revolution built on lessons from budding yeast. Nat. Rev. Genet. *3*, 918-930.

Berry,D.B. and Gasch,A.P. (2008). Stress-activated genomic expression changes serve a preparative role for impending stress in yeast. Mol. Biol. Cell *19*, 4580-4587.

Beuning,P.J. and Musier-Forsyth,K. (1999). Transfer RNA recognition by aminoacyl-tRNA synthetases. Biopolymers *52*, 1-28.

Bjork,G.R., Durand,J.M., Hagervall,T.G., Leipuviene,R., Lundgren,H.K., Nilsson,K., Chen,P., Qian,Q., and Urbonavicius,J. (1999). Transfer RNA modification: influence on translational frameshifting and metabolism. FEBS Lett. *452*, 47-51.

Bjork,G.R., Jacobsson,K., Nilsson,K., Johansson,M.J., Bystrom,A.S., and Persson,O.P. (2001). A primordial tRNA modification required for the evolution of life? EMBO J. *20*, 231-239.

Blight,S.K., Larue,R.C., Mahapatra,A., Longstaff,D.G., Chang,E., Zhao,G., Kang,P.T., Green-Church,K.B., Chan,M.K., and Krzycki,J.A. (2004). Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. Nature *431*, 333-335.

Bochner,B., Gomez,V., Ziman,M., Yang,S., and Brown,S.D. (2010). Phenotype microarray profiling of Zymomonas mobilis ZM4. Appl. Biochem. Biotechnol. *161*, 116-123.

Bochner,B.R. (2003). New technologies to assess genotype-phenotype relationships. Nat. Rev. Genet. *4*, 309-314.

Bochner,B.R. (2009). Global phenotypic characterization of bacteria. FEMS Microbiol. Rev. *33*, 191-205.

Bochner,B.R., Gadzinski,P., and Panomitros,E. (2001). Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome Res. *11*, 1246-1255.

Bock,A., Forchhammer,K., Heider,J., Leinfelder,W., Sawers,G., Veprek,B., and Zinoni,F. (1991). Selenocysteine: the 21st amino acid. Mol. Microbiol. *5*, 515-520.

Boer,V.M., Amini,S., and Botstein,D. (2008). Influence of genotype and nutrition on survival and metabolism of starving yeast. Proc. Natl. Acad. Sci. U. S. A *105*, 6930-6935.

Borovinskaya,M.A., Shoji,S., Fredrick,K., and Cate,J.H. (2008). Structural basis for hygromycin B inhibition of protein biosynthesis. RNA. *14*, 1590-1599.

Bouchonville,K., Forche,A., Tang,K.E., Selmecki,A., and Berman,J. (2009). Aneuploid chromosomes are highly unstable during DNA transformation of Candida albicans. Eukaryot. Cell *8*, 1554-1566.

Braun,B.R. *et al.* (2005). A human-curated annotation of the Candida albicans genome. PLoS. Genet. *1*, 36-57.

Breitschopf,K., Achsel,T., Busch,K., and Gross,H.J. (1995). Identity elements of human tRNA(Leu): structural requirements for converting human tRNA(Ser) into a leucine acceptor in vitro. Nucleic Acids Res. *23*, 3633-3637.

Brown,D.A. and London,E. (1998). Structure and origin of ordered lipid domains in biological membranes. J. Membr. Biol. *164*, 103-114.

Budisa,N., Minks,C., Alefelder,S., Wenger,W., Dong,F., Moroder,L., and Huber,R. (1999). Toward the experimental codon reassignment in vivo: protein building with an expanded amino acid repertoire. FASEB J. *13*, 41-51.

Bukau,B., Weissman,J., and Horwich,A. (2006). Molecular chaperones and protein quality control. Cell *125*, 443-451.

Butler,G. *et al.* (2009). Evolution of pathogenicity and sexual reproduction in eight Candida genomes. Nature *459*, 657-662.

Calderone,R.A. and Fonzi,W.A. (2001). Virulence factors of Candida albicans. Trends Microbiol. *9*, 327-335.

Cannon,R.D., Lamping,E., Holmes,A.R., Niimi,K., Tanabe,K., Niimi,M., and Monk,B.C. (2007). Candida albicans drug resistance another way to cope with stress. Microbiology *153*, 3211-3217.

Castresana,J., Feldmaier-Fuchs,G., and Paabo,S. (1998). Codon reassignment and amino acid composition in hemichordate mitochondria. Proc. Natl. Acad. Sci. U. S. A *95*, 3703-3707.

Causton,H.C., Ren,B., Koh,S.S., Harbison,C.T., Kanin,E., Jennings,E.G., Lee,T.I., True,H.L., Lander,E.S., and Young,R.A. (2001). Remodeling of yeast genome expression in response to environmental changes. Mol. Biol. Cell *12*, 323-337.

Chapeville,F., Lipmann,F., Von-Ehrenstein,G., Weisblum,B., Ray,W.J.Jr., and Benzer,S. (1962). On the role of soluble ribonucleic acid in coding for amino acid. Proc. Natl. Acad. Sci. U. S. A. *48*, 1086-1092.

Chen,D., Toone,W.M., Mata,J., Lyne,R., Burns,G., Kivinen,K., Brazma,A., Jones,N., and Bahler,J. (2003). Global transcriptional responses of fission yeast to environmental stress. Mol. Biol. Cell *14*, 214-229.

Chong,Y.E., Yang,X.L., and Schimmel,P. (2008). Natural homolog of tRNA synthetase editing domain rescues conditional lethality caused by mistranslation. J. Biol. Chem. *283*, 30073-30078.

Cormack,B.P., Bertram,G., Egerton,M., Gow,N.A., Falkow,S., and Brown,A.J. (1997). Yeast-enhanced green fluorescent protein (yEGFP): a reporter of gene expression in Candida albicans. Microbiology *143 (Pt 2)*, 303-311.

Coste,A., Turner,V., Ischer,F., Morschhauser,J., Forche,A., Selmecki,A., Berman,J., Bille,J., and Sanglard,D. (2006). A mutation in Tac1p, a transcription factor regulating CDR1 and CDR2, is coupled with loss of heterozygosity at chromosome 5 to mediate antifungal resistance in Candida albicans. Genetics *172*, 2139-2156.

Crick,F.H. (1958). On protein synthesis. Symp. Soc. Exp. Biol. *12*, 138-163.

Crick,F.H. (1968). The origin of the genetic code. J. Mol. Biol. *38*, 367-379.

Cruz,M.C., Goldstein,A.L., Blankenship,J.R., Del,P.M., Davis,D., Cardenas,M.E., Perfect,J.R., McCusker,J.H., and Heitman,J. (2002). Calcineurin is essential for survival during membrane stress in Candida albicans. EMBO J. *21*, 546-559.

Cusack,S. (1997). Aminoacyl-tRNA synthetases. Curr. Opin. Struct. Biol. *7*, 881-889.

Cusack,S., Berthet-Colominas,C., Hartlein,M., Nassar,N., and Leberman,R. (1990). A second class of synthetase structure revealed by X-ray analysis of Escherichia coli seryl-tRNA synthetase at 2.5 A. Nature *347*, 249-255.

d'Enfert,C. *et al.* (2005). CandidaDB: a genome database for Candida albicans pathogenomics. Nucleic Acids Res. *33*, D353-D357.

Dabas,N. and Morschhauser,J. (2007). Control of ammonium permease expression and filamentous growth by the GATA transcription factors GLN3 and GAT1 in Candida albicans. Eukaryot. Cell *6*, 875-888.

Daniels,K.J., Srikantha,T., Lockhart,S.R., Pujol,C., and Soll,D.R. (2006). Opaque cells signal white cells to form biofilms in Candida albicans. EMBO J. *25*, 2240-2252.

Davis,D. (2003). Adaptation to environmental pH in Candida albicans and its relation to pathogenesis. Curr. Genet. *44*, 1-7.

De Pouplana,L.R. and Schimmel,P. (2004). Aminoacylations of TRNAs: Record-Keepers for the Genetic Code. In: Protein synthesis and ribosome structure, ed. Nierhaus K.H. and Wilson D.N.Weinheim: Wiley-VCH, 169-184.

Dignard,D. and Whiteway,M. (2006). SST2, a regulator of G-protein signaling for the Candida albicans mating response pathway. Eukaryot. Cell *5*, 192-202.

Dirheimer,G., Baranowski,W., and Keith,G. (1995). Variations in tRNA modifications, particularly of their queuine content in higher eukaryotes. Its relation to malignancy grading. Biochimie *77*, 99-103.

Doedt,T., Krishnamurthy,S., Bockmuhl,D.P., Tebarth,B., Stempel,C., Russell,C.L., Brown,A.J., and Ernst,J.F. (2004). APSES proteins regulate morphogenesis and metabolism in Candida albicans. Mol. Biol. Cell *15*, 3167-3180.

Doolittle,R.F. and Handy,J. (1998). Evolutionary anomalies among the aminoacyl-tRNA synthetases. Curr. Opin. Genet. Dev. *8*, 630-636.

Dorazi,R., Lingutla,J.J., and Humayun,M.Z. (2002). Expression of mutant alanine tRNAs increases spontaneous mutagenesis in Escherichia coli. Mol. Microbiol. *44*, 131-141.

Drotschmann,K., Shcherbakova,P.V., and Kunkel,T.A. (2000). Mutator phenotype due to loss of heterozygosity in diploid yeast strains with mutations in MSH2 and MLH1. Toxicol. Lett. *112-113*, 239-244.

Dunham,M.J., Badrane,H., Ferea,T., Adams,J., Brown,P.O., Rosenzweig,F., and Botstein,D. (2002). Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A *99*, 16144-16149.

Effraim,P.R., Wang,J., Englander,M.T., Avins,J., Leyh,T.S., Gonzalez,R.L., Jr., and Cornish,V.W. (2009). Natural amino acids do not require their native tRNAs for efficient selection by the ribosome. Nat. Chem. Biol. *5*, 947-953.

Eisen,M.B., Spellman,P.T., Brown,P.O., and Botstein,D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U. S. A *95*, 14863-14868.

Enjalbert,B., Nantel,A., and Whiteway,M. (2003). Stress-induced gene expression in Candida albicans: absence of a general stress response. Mol. Biol. Cell *14*, 1460-1467.

Enjalbert,B., Smith,D.A., Cornell,M.J., Alam,I., Nicholls,S., Brown,A.J., and Quinn,J. (2006). Role of the Hog1 stress-activated protein kinase in the global transcriptional response to stress in the fungal pathogen Candida albicans. Mol. Biol. Cell *17*, 1018-1032.

Eriani,G., Delarue,M., Poch,O., Gangloff,J., and Moras,D. (1990). Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature *347*, 203-206.

Estruch,F. (2000). Stress-controlled transcription factors, stress-induced genes and stress tolerance in budding yeast. FEMS Microbiol. Rev. *24*, 469-486.

Farabaugh,P.J. (1996). Programmed translational frameshifting. Microbiol. Rev. *60*, 103-134.

Farabaugh,P.J. and Bjork,G.R. (1999). How translational accuracy influences reading frame maintenance. EMBO J. *18*, 1427-1434.

Felts,S.J., Karnitz,L.M., and Toft,D.O. (2007). Functioning of the Hsp90 machine in chaperoning checkpoint kinase I (Chk1) and the progesterone receptor (PR). Cell Stress. Chaperones. *12*, 353-363.

Feng,L., Sheppard,K., Tumbula-Hansen,D., and Soll,D. (2005a). Gln-tRNAGln formation from Glu-tRNAGln requires cooperation of an asparaginase and a Glu-tRNAGln kinase. J. Biol. Chem. *280*, 8150-8155.

Feng,L., Yuan,J., Toogood,H., Tumbula-Hansen,D., and Soll,D. (2005b). Aspartyl-tRNA synthetase requires a conserved proline in the anticodon-binding loop for tRNA(Asn) recognition in vivo. J. Biol. Chem. *280*, 20638-20641.

Finley,K.R., Bouchonville,K.J., Quick,A., and Berman,J. (2008). Dynein-dependent nuclear dynamics affect morphogenesis in Candida albicans by means of the Bub2p spindle checkpoint. J. Cell Sci. *121*, 466-476.

First,E.A. and Fersht,A.R. (1993). Mutational and kinetic analysis of a mobile loop in tyrosyl-tRNA synthetase. Biochemistry *32*, 13658-13663.

Forche,A., Abbey,D., Pisithkul,T., Weinzierl,M.A., Ringstrom,T., Bruck,D., Petersen,K., and Berman,J. (2011). Stress alters rates and types of loss of heterozygosity in Candida albicans. MBio. *2*.

Forche,A., Alby,K., Schaefer,D., Johnson,A.D., Berman,J., and Bennett,R.J. (2008). The parasexual cycle in Candida albicans provides an alternative pathway to meiosis for the formation of recombinant strains. PLoS. Biol. *6*, e110.

Frohner,I.E., Bourgeois,C., Yatsyk,K., Majer,O., and Kuchler,K. (2009). Candida albicans cell surface superoxide dismutases degrade host-derived reactive oxygen species to escape innate immune surveillance. Mol. Microbiol. *71*, 240-252.

Gasch,A.P. (2007). Comparative genomics of the environmental stress response in ascomycete fungi. Yeast *24*, 961-976.

Gasch,A.P., Moses,A.M., Chiang,D.Y., Fraser,H.B., Berardini,M., and Eisen,M.B. (2004). Conservation and evolution of cis-regulatory systems in ascomycete fungi. PLoS. Biol. *2*, e398.

Gasch,A.P., Spellman,P.T., Kao,C.M., Carmel-Harel,O., Eisen,M.B., Storz,G., Botstein,D., and Brown,P.O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell *11*, 4241-4257.

Gasch,A.P. and Werner-Washburne,M. (2002). The genomics of yeast responses to environmental stress and starvation. Funct. Integr. Genomics *2*, 181-192.

Gaston,M.A., Jiang,R., and Krzycki,J.A. (2011a). Functional context, biosynthesis, and genetic encoding of pyrrolysine. Curr. Opin. Microbiol. *14*, 342-349.

Gaston,M.A., Zhang,L., Green-Church,K.B., and Krzycki,J.A. (2011b). The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. Nature *471*, 647-650.

Gentleman,R.C. *et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

Gesteland,R.F. and Atkins,J.F. (1996). Recoding: dynamic reprogramming of translation. Annu. Rev. Biochem. *65*, 741-768.

Ghannoum,M.A. and Rice,L.B. (1999). Antifungal agents: mode of action, mechanisms of resistance, and correlation of these mechanisms with bacterial resistance. Clin. Microbiol. Rev. *12*, 501-517.

Giege,R. (2006). The early history of tRNA recognition by aminoacyl-tRNA synthetases. J. Biosci. *31*, 477-488.

Giege,R. (2008). Toward a more complete view of tRNA biology. Nat. Struct. Mol. Biol. *15*, 1007-1014.

Giege,R., Sissler,M., and Florentz,C. (1998). Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res. *26*, 5017-5035.

Glowacki,R., Bald,E., and Jakubowski,H. (2010). Identification and origin of Nepsilon-homocysteinyl-lysine isopeptide in humans and mice. Amino. Acids *39*, 1563-1569.

Gomes,A.C., Miranda,I., Silva,R.M., Moura,G.R., Thomas,B., Akoulitchev,A., and Santos,M.A. (2007). A genetic code alteration generates a proteome of high diversity in the human pathogen Candida albicans. Genome Biol. *8*, R206.

Goodenbour,J.M. and Pan,T. (2006). Diversity of tRNA genes in eukaryotes. Nucleic Acids Res. *34*, 6137-6146.

Goyal,S. and Khuller,G.K. (1994). Structural and functional role of lipids in yeast and mycelial forms of Candida albicans. Lipids *29*, 793-797.

Graser,Y., Volovsek,M., Arrington,J., Schonian,G., Presber,W., Mitchell,T.G., and Vilgalys,R. (1996). Molecular markers reveal that population structure of the human pathogen Candida albicans exhibits both clonality and recombination. Proc. Natl. Acad. Sci. U. S. A *93*, 12473-12477.

Gromadski,K.B. and Rodnina,M.V. (2004). Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. Mol. Cell *13*, 191-200.

Guo,M. and Schimmel,P. (2011). Structural analyses clarify the complex control of mistranslation by tRNA synthetases. Curr. Opin. Struct. Biol.

Haase,S.B. and Lew,D.J. (1997). Flow cytometric analysis of DNA content in budding yeast. Methods Enzymol. *283*, 322-332.

Halbeisen,R.E. and Gerber,A.P. (2009). Stress-Dependent Coordination of Transcriptome and Translatome in Yeast. PLoS. Biol. *7*, e105.

Hanyu,N., Kuchino,Y., Nishimura,S., and Beier,H. (1986). Dramatic events in ciliate evolution: alteration of UAA and UAG termination codons to glutamine codons due to anticodon mutations in two Tetrahymena tRNAs. EMBO J. *5*, 1307-1311.

Hatfield,D. and Diamond,A. (1993). UGA: a split personality in the universal genetic code. Trends Genet. *9*, 69-70.

Heitzler,J., Marechal-Drouard,L., Dirheimer,G., and Keith,G. (1992). Use of a dot blot hybridization method for identification of pure tRNA species on different membranes. Biochim. Biophys. Acta *1129*, 273-277.

Henson,J., Tischler,G., and Ning,Z. (2012). Next-generation sequencing and large genome assemblies. Pharmacogenomics. *13*, 901-915.

Herr,A.J., Gesteland,R.F., and Atkins,J.F. (2000). One protein from two open reading frames: mechanism of a 50 nt translational bypass. EMBO J. *19*, 2671-2680.

Herrero,E., Ros,J., Belli,G., and Cabiscol,E. (2008). Redox control and oxidative stress in yeast cells. Biochim. Biophys. Acta *1780*, 1217-1235.

Hilton,C., Markie,D., Corner,B., Rikkerink,E., and Poulter,R. (1985). Heat shock induces chromosome loss in the yeast Candida albicans. Mol. Gen. Genet. *200*, 162-168.

Hoesl,M.G. and Budisa,N. (2012). Recent advances in genetic code engineering in Escherichia coli. Curr. Opin. Biotechnol.

Hogues,H., Lavoie,H., Sellam,A., Mangos,M., Roemer,T., Purisima,E., Nantel,A., and Whiteway,M. (2008). Transcription factor substitution during the evolution of fungal ribosome regulation. Mol. Cell *29*, 552-562.

Holley,R.W., EVERETT,G.A., MADISON,J.T., and ZAMIR,A. (1965). Nucleotide sequences in the yeast alanine transfer ribonucleic acid. J. Biol. Chem. *240*, 2122-2128.

Homann,O.R., Dea,J., Noble,S.M., and Johnson,A.D. (2009). A phenotypic profile of the Candida albicans regulatory network. PLoS. Genet. *5*, e1000783.

Hoot,S.J., Zheng,X., Potenski,C.J., White,T.C., and Klein,H.L. (2011). The role of Candida albicans homologous recombination factors Rad54 and Rdh54 in DNA damage sensitivity. BMC. Microbiol. *11*, 214.

Hromatka,B.S., Noble,S.M., and Johnson,A.D. (2005). Transcriptional response of Candida albicans to nitric oxide and the role of the YHB1 gene in nitrosative stress and virulence. Mol. Biol. Cell *16*, 4814-4826.

Hube,B. (2004). From commensal to pathogen: stage- and tissue-specific gene expression of Candida albicans. Curr. Opin. Microbiol. *7*, 336-341.

Hughes,T.R. *et al.* (2000a). Functional discovery via a compendium of expression profiles. Cell *102*, 109-126.

Hughes,T.R. *et al.* (2000b). Widespread aneuploidy revealed by DNA microarray expression profiling. Nat. Genet. *25*, 333-337.

Hull,C.M. and Johnson,A.D. (1999). Identification of a mating type-like locus in the asexual pathogenic yeast Candida albicans. Science *285*, 1271-1275.

Ibba,M., Curnow,A.W., and Soll,D. (1997a). Aminoacyl-tRNA synthesis: divergent routes to a common goal. Trends Biochem. Sci. *22*, 39-42.

Ibba,M., Morgan,S., Curnow,A.W., Pridmore,D.R., Vothknecht,U.C., Gardner,W., Lin,W., Woese,C.R., and Soll,D. (1997b). A euryarchaeal lysyl-tRNA synthetase: resemblance to class I synthetases. Science *278*, 1119-1122.

Ibba,M. and Soll,D. (1999). Quality control mechanisms during translation. Science *286*, 1893-1897.

Ibba,M. and Soll,D. (2004). Aminoacyl-tRNAs: setting the limits of the genetic code. Genes Dev. *18*, 731-738.

Ihmels,J., Bergmann,S., and Barkai,N. (2004). Defining transcription modules using large-scale gene expression data. Bioinformatics. *20*, 1993-2003.

Ikemura,T. (1982). Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. J. Mol. Biol. *158*, 573-597.

Inagaki,Y. and Doolittle,W.F. (2001). Class I release factors in ciliates with variant genetic codes. Nucleic Acids Res. *29*, 921-927.

Iwaguchi,S., Homma,M., and Tanaka,K. (1992). Clonal variation of chromosome size derived from the rDNA cluster region in Candida albicans. J. Gen. Microbiol. *138*, 1177-1184.

Iwaguchi,S.I., Sato,M., Magee,B.B., Magee,P.T., Makimura,K., and Suzuki,T. (2001). Extensive chromosome translocation in a clinical isolate showing the distinctive carbohydrate assimilation profile from a candidiasis patient. Yeast *18*, 1035-1046.

Jakubowski,H. (2008). The pathophysiological hypothesis of homocysteine thiolactone-mediated vascular disease. J. Physiol Pharmacol. *59 Suppl 9*, 155-167.

Jakubowski,H. (2011). Quality control in tRNA charging. Wiley. Interdiscip. Rev. RNA.

Jones,T. *et al.* (2004). The diploid genome sequence of Candida albicans. Proc. Natl. Acad. Sci. U. S. A *101*, 7329-7334.

Jorgensen,P., Rupes,I., Sharom,J.R., Schneper,L., Broach,J.R., and Tyers,M. (2004). A dynamic transcriptional network communicates growth potential to ribosome synthesis and critical cell size. Genes Dev. *18*, 2491-2505.

Kabir,M.A., Ahmad,A., Greenberg,J.R., Wang,Y.K., and Rustchenko,E. (2005). Loss and gain of chromosome 5 controls growth of Candida albicans on sorbose due to dispersed redundant negative regulators. Proc. Natl. Acad. Sci. U. S. A *102*, 12147-12152.

Kahvejian,A., Quackenbush,J., and Thompson,J.F. (2008). What would you do if you could sequence everything? Nat. Biotechnol. *26*, 1125-1133.

Kapp,L.D., Kolitz,S.E., and Lorsch,J.R. (2006). Yeast initiator tRNA identity elements cooperate to influence multiple steps of translation initiation. RNA. *12*, 751-764.

Kapp,L.D. and Lorsch,J.R. (2004). The molecular mechanics of eukaryotic translation. Annu. Rev. Biochem. *73*, 657-704.

Karababa,M., Coste,A.T., Rognon,B., Bille,J., and Sanglard,D. (2004). Comparison of gene expression profiles of Candida albicans azole-resistant clinical isolates and laboratory strains exposed to drugs inducing multidrug transporters
2. Antimicrob. Agents Chemother. *48*, 3064-3079.

Khaled H., Abu-Elteen, and Mawieh Hamad (2011). Antifungal Agents for Use in Human Therapy. In: Fungi: Biology and Applications, ed. Kevin Kavanagh John Wiley & Sons, Ltd., 279-311.

Kim,J. and Sudbery,P. (2011). Candida albicans, a major human fungal pathogen. J. Microbiol. *49*, 171-177.

Klann,E. and Dever,T.E. (2004). Biochemical mechanisms for translational regulation in synaptic plasticity. Nat. Rev. Neurosci. *5*, 931-942.

Knight,R.D., Freeland,S.J., and Landweber,L.F. (2001). Rewiring the keyboard: evolvability of the genetic code. Nat. Rev. Genet. *2*, 49-58.

Koonin,E.V. and Novozhilov,A.S. (2009). Origin and evolution of the genetic code: the universal enigma. IUBMB. Life *61*, 99-111.

Kramer,E.B. and Farabaugh,P.J. (2007). The frequency of translational misreading errors in E. coli is largely determined by tRNA competition. RNA. *13*, 87-96.

Krzycki,J.A. (2005). The direct genetic encoding of pyrrolysine. Curr. Opin. Microbiol. *8*, 706-712.

Kusch,H., Engelmann,S., Albrecht,D., Morschhauser,J., and Hecker,M. (2007). Proteomic analysis of the oxidative stress response in Candida albicans. Proteomics. *7*, 686-697.

Lackner,D.H. and Bahler,J. (2008). Translational control of gene expression from transcripts to transcriptomes. Int. Rev. Cell Mol. Biol. *271*, 199-251.

Lancaster,L., Kiel,M.C., Kaji,A., and Noller,H.F. (2002). Orientation of ribosome recycling factor in the ribosome from directed hydroxyl radical probing. Cell *111*, 129-140.

Lange,B.M., Bachi,A., Wilm,M., and Gonzalez,C. (2000). Hsp90 is a core centrosomal component and is required at different stages of the centrosome cycle in Drosophila and vertebrates. EMBO J. *19*, 1252-1262.

Lavoie,H., Hogues,H., and Whiteway,M. (2009). Rearrangements of the transcriptional regulatory networks of metabolic pathways in fungi. Curr. Opin. Microbiol. *12*, 655-663.

Leach,M.D., Stead,D.A., Argo,E., and Brown,A.J. (2011). Identification of sumoylation targets, combined with inactivation of SMT3, reveals the impact of sumoylation upon growth, morphology, and stress resistance in the pathogen Candida albicans
4. Mol. Biol. Cell *22*, 687-702.

Lee,J.W., Beebe,K., Nangle,L.A., Jang,J., Longo-Guess,C.M., Cook,S.A., Davisson,M.T., Sundberg,J.P., Schimmel,P., and Ackerman,S.L. (2006). Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature *443*, 50-55.

Legrand,M., Chan,C.L., Jauert,P.A., and Kirkpatrick,D.T. (2007). Role of DNA mismatch repair and double-strand break repair in genome stability and antifungal drug resistance in Candida albicans. Eukaryot. Cell *6*, 2194-2205.

Legrand,M., Chan,C.L., Jauert,P.A., and Kirkpatrick,D.T. (2011). The contribution of the S-phase checkpoint genes MEC1 and SGS1 to genome stability maintenance in Candida albicans. Fungal. Genet. Biol. *48*, 823-830.

Legrand,M., Forche,A., Selmecki,A., Chan,C., Kirkpatrick,D.T., and Berman,J. (2008). Haplotype mapping of a diploid non-meiotic organism using existing and induced aneuploidies. PLoS. Genet. *4*, e1.

Lephart,P.R., Chibana,H., and Magee,P.T. (2005). Effect of the major repeat sequence on chromosome loss in Candida albicans. Eukaryot. Cell *4*, 733-741.

Lephart,P.R. and Magee,P.T. (2006). Effect of the major repeat sequence on mitotic recombination in Candida albicans. Genetics *174*, 1737-1744.

Levy,S., Ihmels,J., Carmi,M., Weinberger,A., Friedlander,G., and Barkai,N. (2007). Strategy of transcription regulation in the budding yeast. PLoS. One. *2*, e250.

Li,H. and Durbin,R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. *26*, 589-595.

Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., and Durbin,R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics. *25*, 2078-2079.

Li,L., Boniecki,M.T., Jaffe,J.D., Imai,B.S., Yau,P.M., Luthey-Schulten,Z.A., and Martinis,S.A. (2011). Naturally occurring aminoacyl-tRNA synthetases editing-domain mutations that cause mistranslation in Mycoplasma parasites. Proc. Natl. Acad. Sci. U. S. A *108*, 9378-9383.

Li,M. and Tzagoloff,A. (1979). Assembly of the mitochondrial membrane system: sequences of yeast mitochondrial valine and an unusual threonine tRNA gene. Cell *18*, 47-53.

Lindquist,S. and Craig,E.A. (1988). The heat-shock proteins. Annu. Rev. Genet. *22*, 631-677.

Ling,J., Reynolds,N., and Ibba,M. (2009). Aminoacyl-tRNA synthesis and translational quality control. Annu. Rev. Microbiol. *63*, 61-78.

Link,A.J., Mock,M.L., and Tirrell,D.A. (2003). Non-canonical amino acids in protein engineering. Curr. Opin. Biotechnol. *14*, 603-609.

Liti,G. *et al.* (2009). Population genomics of domestic and wild yeasts. Nature *458*, 337-341.

Liu,C.C. and Schultz,P.G. (2010). Adding new chemistries to the genetic code. Annu. Rev. Biochem. *79*, 413-444.

Lockhart,S.R., Daniels,K.J., Zhao,R., Wessels,D., and Soll,D.R. (2003). Cell biology of mating in Candida albicans. Eukaryot. Cell *2*, 49-61.

Lorenz,M.C., Bender,J.A., and Fink,G.R. (2004). Transcriptional response of Candida albicans upon internalization by macrophages. Eukaryot. Cell *3*, 1076-1087.

Lovett,P.S., Ambulos,N.P., Jr., Mulbry,W., Noguchi,N., and Rogers,E.J. (1991). UGA can be decoded as tryptophan at low efficiency in Bacillus subtilis. J. Bacteriol. *173*, 1810-1812.

Lozupone,C.A., Knight,R.D., and Landweber,L.F. (2001). The molecular basis of nuclear genetic code change in ciliates. Curr. Biol. *11*, 65-74.

Lynch,M. *et al.* (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc. Natl. Acad. Sci. U. S. A *105*, 9272-9277.

Lyons,C.N. and White,T.C. (2000). Transcriptional analyses of antifungal drug resistance in Candida albicans. Antimicrob. Agents Chemother. *44*, 2296-2303.

MacCallum,D.M., Coste,A., Ischer,F., Jacobsen,M.D., Odds,F.C., and Sanglard,D. (2010). Genetic dissection of azole resistance mechanisms in Candida albicans and their validation in a mouse model of disseminated infection. Antimicrob. Agents Chemother. *54*, 1476-1483.

Macino,G., Coruzzi,G., Nobrega,F.G., Li,M., and Tzagoloff,A. (1979). Use of the UGA terminator as a tryptophan codon in yeast mitochondria. Proc. Natl. Acad. Sci. U. S. A *76*, 3784-3785.

Magee,B.B. and Magee,P.T. (2000). Induction of mating in Candida albicans by construction of MTLa and MTLalpha strains. Science *289*, 310-313.

Magee,P.T. (2007). Genome Structure and Dynamics in Candida Albicans. In: Candida: comparative and functional genomics, ed. C.d'Enfert and B.HubeNorfolk, United Kingdom: Caister Academic Press, 7-26.

Magee,P.T., Bowdin,L., and Staudinger,J. (1992). Comparison of molecular typing methods for Candida albicans. J. Clin. Microbiol. *30*, 2674-2679.

Mahlab,S., Tuller,T., and Linial,M. (2012). Conservation of the relative tRNA composition in healthy and cancerous tissues. RNA.

Márquez V. and Nierhaus K.H. (2004). TRNA and Synthetases. In: Protein synthesis and ribosome structure, ed. K.H.Nierhaus and D.N.WilsonWeinheim: WILEY-VCH Verlag GmbH & Co. KGaA, 145-184.

Martinis,S.A. and Boniecki,M.T. (2010). The balance between pre- and post-transfer editing in tRNA synthetases. FEBS Lett. *584*, 455-459.

Martinis,S.A., Plateau,P., Cavarelli,J., and Florentz,C. (1999a). Aminoacyl-tRNA synthetases: a family of expanding functions. Mittelwihr, France, October 10-15, 1999. EMBO J. *18*, 4591-4596.

Martinis,S.A., Plateau,P., Cavarelli,J., and Florentz,C. (1999b). Aminoacyl-tRNA synthetases: a new image for a classical family. Biochimie *81*, 683-700.

Massey,S.E., Moura,G., Beltrao,P., Almeida,R., Garey,J.R., Tuite,M.F., and Santos,M.A. (2003). Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in Candida spp. Genome Res. *13*, 544-557.

Mazumder,B., Seshadri,V., and Fox,P.L. (2003). Translational control by the 3'-UTR: the ends specify the means. Trends Biochem. Sci. *28*, 91-98.

McClellan,A.J., Tam,S., Kaganovich,D., and Frydman,J. (2005). Protein quality control: chaperones culling corrupt conformations. Nat. Cell Biol. *7*, 736-741.

Meyer,F., Schmidt,H.J., Plumper,E., Hasilik,A., Mersmann,G., Meyer,H.E., Engstrom,A., and Heckmann,K. (1991). UGA is translated as cysteine in pheromone 3 of Euplotes octocarinatus. Proc. Natl. Acad. Sci. U. S. A *88*, 3758-3761.

Meyerovich,M., Mamou,G., and Ben-Yehuda,S. (2010). Visualizing high error levels during gene expression in living bacterial cells. Proc. Natl. Acad. Sci. U. S. A *107*, 11543-11548.

Miranda,I., Rocha,R., Santos,M.C., Mateus,D.D., Moura,G.R., Carreto,L., and Santos,M.A. (2007). A genetic code alteration is a phenotype diversity generator in the human pathogen Candida albicans. PLoS. One. *2*, e996.

Miranda,I., Silva,R., and Santos,M.A. (2006). Evolution of the genetic code in yeasts. Yeast *23*, 203-213.

Moore,P.B. and Steitz,T.A. (2003). After the ribosome structures: how does peptidyl transferase work? RNA. *9*, 155-159.

Moura,G., Pinheiro,M., Silva,R., Miranda,I., Afreixo,V., Dias,G., Freitas,A., Oliveira,J.L., and Santos,M.A. (2005). Comparative context analysis of codon pairs on an ORFeome scale. Genome Biol. *6*, R28.

Moura,G.R., Carreto,L.C., and Santos,M.A. (2009). Genetic code ambiguity: an unexpected source of proteome innovation and phenotypic diversity. Curr. Opin. Microbiol. *12*, 631-637.

Mukai,T., Hayashi,A., Iraha,F., Sato,A., Ohtake,K., Yokoyama,S., and Sakamoto,K. (2010). Codon reassignment in the Escherichia coli genetic code. Nucleic Acids Res. *38*, 8188-8195.

Murad,A.M., Lee,P.R., Broadbent,I.D., Barelle,C.J., and Brown,A.J. (2000). Clp10, an efficient and convenient integrating vector for Candida albicans. Yeast *16*, 325-327.

Muramatsu,T., Nishikawa,K., Nemoto,F., Kuchino,Y., Nishimura,S., Miyazawa,T., and Yokoyama,S. (1988). Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. Nature *336*, 179-181.

Murphy,H.S. and Humayun,M.Z. (1997). Escherichia coli cells expressing a mutant glyV (glycine tRNA) gene have a UVM-constitutive phenotype: implications for mechanisms underlying the mutA or mutC mutator effect. J. Bacteriol. *179*, 7507-7514.

Nangle,L.A., Motta,C.M., and Schimmel,P. (2006). Global effects of mistranslation from an editing defect in mammalian cells. Chem. Biol. *13*, 1091-1100.

Nantel,A. (2006). The long hard road to a completed Candida albicans genome. Fungal. Genet. Biol. *43*, 311-315.

Nett,J.E., Lepak,A.J., Marchillo,K., and Andes,D.R. (2009). Time course global gene expression analysis of an in vivo Candida biofilm. J. Infect. Dis. *200*, 307-313.

Netzer,N. *et al.* (2009). Innate immune and chemically triggered oxidative stress modifies translational fidelity. Nature *462*, 522-526.

Neuman,J.A., Isakov,O., and Shomron,N. (2012). Analysis of insertion-deletion from deep-sequencing data: software evaluation for optimal detection. Brief. Bioinform.

Newbury,S.F. (2006). Control of mRNA stability in eukaryotes. Biochem. Soc. Trans. *34*, 30-34.

Ng,S.B. *et al.* (2010). Exome sequencing identifies the cause of a mendelian disorder. Nat. Genet. *42*, 30-35.

Ng,S.B. *et al.* (2009). Targeted capture and massively parallel sequencing of 12 human exomes. Nature *461*, 272-276.

Nicholls,S., Straffon,M., Enjalbert,B., Nantel,A., Macaskill,S., Whiteway,M., and Brown,A.J. (2004). Msn2- and Msn4-like transcription factors play no obvious roles in the stress responses of the fungal pathogen Candida albicans. Eukaryot. Cell *3*, 1111-1123.

Nierhaus K.H. and Wilson D.N. (2004). Protein Synthesis and Ribosome Structure, Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA.

Nirenberg,M. *et al.* (1966). The RNA code and protein synthesis. Cold Spring Harb. Symp. Quant. Biol. *31*, 11-24.

Nirenberg,M. and Leder,P. (1964). RNA codewords and protein synthesis. The effect of trinucleotides upon the binding of sRNA to ribosomes. Science *145*, 1399-1407.

Noble,S.M. and Johnson,A.D. (2005). Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen Candida albicans. Eukaryot. Cell *4*, 298-309.

Novozhilov,A.S., Wolf,Y.I., and Koonin,E.V. (2007). Evolution of the genetic code: partial optimization of a random code for robustness to translation error in a rugged fitness landscape. Biol. Direct. *2*, 24.

O'Sullivan,J.M., Davenport,J.B., and Tuite,M.F. (2001a). Codon reassignment and the evolving genetic code: problems and pitfalls in post-genome analysis. Trends Genet. *17*, 20-22.

O'Sullivan,J.M., Mihr,M.J., Santos,M.A., and Tuite,M.F. (2001b). Seryl-tRNA synthetase is not responsible for the evolution of CUG codon reassignment in Candida albicans. Yeast *18*, 313-322.

O'Sullivan,J.M., Mihr,M.J., Santos,M.A., and Tuite,M.F. (2001c). The Candida albicans gene encoding the cytoplasmic leucyl-tRNA synthetase: implications for the evolution of CUG codon reassignment. Gene *275*, 133-140.

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

Oba,T., Andachi,Y., Muto,A., and Osawa,S. (1991b). Translation in vitro of codon UGA as tryptophan in Mycoplasma capricolum. Biochimie *73*, 1109-1112.

Odds,F.C. (1994). Pathogenesis of Candida infections. J. Am. Acad. Dermatol. *31*, S2-S5.

Odds,F.C. *et al.* (2007). Molecular phylogenetics of Candida albicans. Eukaryot. Cell *6*, 1041-1052.

Ogle,J.M. and Ramakrishnan,V. (2005). Structural insights into translational fidelity. Annu. Rev. Biochem. *74*, 129-177.

Ohama,T., Muto,A., and Osawa,S. (1990). Role of GC-biased mutation pressure on synonymous codon choice in Micrococcus luteus, a bacterium with a high genomic GC-content. Nucleic Acids Res. *18*, 1565-1569.

Ohama,T., Suzuki,T., Mori,M., Osawa,S., Ueda,T., Watanabe,K., and Nakase,T. (1993). Non-universal decoding of the leucine codon CUG in several Candida species. Nucleic Acids Res. *21*, 4039-4045.

Osawa,S., Collins,D., Ohama,T., Jukes,T.H., and Watanabe,K. (1990). Evolution of the mitochondrial genetic code. III. Reassignment of CUN codons from leucine to threonine during evolution of yeast mitochondria. J. Mol. Evol. *30*, 322-328.

Osawa,S. and Jukes,T.H. (1989). Codon reassignment (codon capture) in evolution. J. Mol. Evol. *28*, 271-278.

Osawa,S., Jukes,T.H., Watanabe,K., and Muto,A. (1992). Recent evidence for evolution of the genetic code. Microbiol. Rev. *56*, 229-264.

Otero,J.M. *et al.* (2010). Whole genome sequencing of Saccharomyces cerevisiae: from genotype to phenotype for improved metabolic engineering applications. BMC. Genomics *11*, 723.

Palmenberg,A.C., Spiro,D., Kuzmickas,R., Wang,S., Djikeng,A., Rathe,J.A., Fraser-Liggett,C.M., and Liggett,S.B. (2009). Sequencing and analyses of all known human rhinovirus genomes reveal structure and evolution. Science *324*, 55-59.

Pape,L.K., Koerner,T.J., and Tzagoloff,A. (1985). Characterization of a yeast nuclear gene (MST1) coding for the mitochondrial threonyl-tRNA1 synthetase. J. Biol. Chem. *260*, 15362-15370.

Pape,T., Wintermeyer,W., and Rodnina,M. (1999). Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. EMBO J. *18*, 3800-3807.

Pelak,K. *et al.* (2010). The characterization of twenty sequenced human genomes. PLoS. Genet. *6*.

Percudani,R., Pavesi,A., and Ottonello,S. (1997). Transfer RNA gene redundancy and translational selection in Saccharomyces cerevisiae. J. Mol. Biol. *268*, 322-330.

Perreau,V.M., Keith,G., Holmes,W.M., Przykorska,A., Santos,M.A., and Tuite,M.F. (1999). The Candida albicans CUG-decoding ser-tRNA has an atypical anticodon stem-loop structure. J. Mol. Biol. *293*, 1039-1053.

Pezo,V., Metzgar,D., Hendrickson,T.L., Waas,W.F., Hazebrouck,S., Doring,V., Marliere,P., Schimmel,P., and De Crecy-Lagard,V. (2004). Artificially ambiguous genetic code confers growth yield advantage. Proc. Natl. Acad. Sci. U. S. A *101*, 8593-8597.

Pickart,C.M. and Cohen,R.E. (2004). Proteasomes and their kin: proteases in the machine age. Nat. Rev. Mol. Cell Biol. *5*, 177-187.

Pina-Vaz,C., Sansonetty,F., Rodrigues,A.G., Costa-Oliveira,S., Tavares,C., and Martinez-de-Oliveira,J. (2001). Cytometric approach for a rapid evaluation of susceptibility of Candida strains to antifungals. Clin. Microbiol. Infect. *7*, 609-618.

Poole,E.S., Major,L.L., Cridge,A.G., and Tate,W.P. (2003). The Mechanism of Recoding in Pro- and Eukaryotes. In: Protein synthesis and ribosome structure, ed. Nierhaus K.H. and Wilson D.N.Weinheim: Wiley-VCH, 397-426.

Ramakrishnan,V. (2002). Ribosome structure and the mechanism of translation. Cell *108*, 557-572.

Raue,H.A. (1994). Metabolic stability of mRNA in yeast--a potential target for modulating productivity? Trends Biotechnol. *12*, 444-449.

Reynolds,N.M., Lazazzera,B.A., and Ibba,M. (2010a). Cellular mechanisms that control mistranslation. Nat. Rev. Microbiol. *8*, 849-856.

Reynolds,N.M., Ling,J., Roy,H., Banerjee,R., Repasky,S.E., Hamel,P., and Ibba,M. (2010b). Cell-specific differences in the requirements for translation quality control. Proc. Natl. Acad. Sci. U. S. A *107*, 4063-4068.

Riggle,P.J. and Kumamoto,C.A. (2000). Role of a Candida albicans P1-type ATPase in resistance to copper and silver ion toxicity
1. J. Bacteriol. *182*, 4899-4905.

Rocha,R., Pereira,P.J., Santos,M.A., and edo-Ribeiro,S. (2011). Unveiling the structural basis for translational ambiguity tolerance in a human fungal pathogen. Proc. Natl. Acad. Sci. U. S. A *108*, 14091-14096.

Rodnina,M.V. (2012). Quality control of mRNA decoding on the bacterial ribosome. Adv. Protein Chem. Struct. Biol. *86*, 95-128.

Rodnina,M.V. and Wintermeyer,W. (2001a). Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms. Annu. Rev. Biochem. *70*, 415-435.

Rodnina,M.V. and Wintermeyer,W. (2001b). Ribosome fidelity: tRNA discrimination, proofreading and induced fit. Trends Biochem. Sci. *26*, 124-130.

Ruan,B., Palioura,S., Sabina,J., Marvin-Guy,L., Kochhar,S., Larossa,R.A., and Soll,D. (2008). Quality control despite mistranslation caused by an ambiguous genetic code. Proc. Natl. Acad. Sci. U. S. A *105*, 16502-16507.

Ruffalo,M., Koyuturk,M., Ray,S., and Laframboise,T. (2012). Accurate estimation of short read mapping quality for next-generation genome sequencing. Bioinformatics. *28*, i349-i355.

Rustchenko,E. (2007). Chromosome instability in Candida albicans. FEMS Yeast Res. *7*, 2-11.

Rustchenko,E.P., Curran,T.M., and Sherman,F. (1993). Variations in the number of ribosomal DNA units in morphological mutants and normal strains of Candida albicans and in normal strains of Saccharomyces cerevisiae. J. Bacteriol. *175*, 7189-7199.

Saeed,A.I., Bhagabati,N.K., Braisted,J.C., Liang,W., Sharov,V., Howe,E.A., Li,J., Thiagarajan,M., White,J.A., and Quackenbush,J. (2006). TM4 microarray software suite. Methods Enzymol. *411*, 134-193.

Sampaio,J.P. and Goncalves,P. (2008). Natural populations of Saccharomyces kudriavzevii in Portugal are associated with oak bark and are sympatric with S. cerevisiae and S. paradoxus. Appl. Environ. Microbiol. *74*, 2144-2152.

Sanchez-Silva,R., Villalobo,E., Morin,L., and Torres,A. (2003). A new noncanonical nuclear genetic code: translation of UAA into glutamate. Curr. Biol. *13*, 442-447.

Sanglard,D. (2002). Resistance of human fungal pathogens to antifungal drugs. Curr. Opin. Microbiol. *5*, 379-385.

Sanglard,D., Ischer,F., Marchetti,O., Entenza,J., and Bille,J. (2003a). Calcineurin A of Candida albicans: involvement in antifungal tolerance, cell morphogenesis and virulence 2. Mol. Microbiol. *48*, 959-976.

Sanglard,D., Ischer,F., Parkinson,T., Falconer,D., and Bille,J. (2003b). Candida albicans mutations in the ergosterol biosynthetic pathway and resistance to several antifungal agents. Antimicrob. Agents Chemother. *47*, 2404-2412.

Santos,M.A., Cheesman,C., Costa,V., Moradas-Ferreira,P., and Tuite,M.F. (1999). Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in Candida spp. Mol. Microbiol. *31*, 937-947.

Santos,M.A., Gomes,A.C., Santos,M.C., Carreto,L.C., and Moura,G.R. (2011). The genetic code of the fungal CTG clade. C. R. Biol. *334*, 607-611.

Santos,M.A., Keith,G., and Tuite,M.F. (1993). Non-standard translational events in Candida albicans mediated by an unusual seryl-tRNA with a 5'-CAG-3' (leucine) anticodon. EMBO J. *12*, 607-616.

Santos,M.A., Moura,G., Massey,S.E., and Tuite,M.F. (2004). Driving change: the evolution of alternative genetic codes. Trends Genet. *20*, 95-102.

Santos,M.A., Perreau,V.M., 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. *15*, 5060-5068.

Santos,M.A. and Tuite,M.F. (1995). The CUG codon is decoded in vivo as serine and not leucine in Candida albicans. Nucleic Acids Res. *23*, 1481-1486.

Santos,M.A., Ueda,T., Watanabe,K., and Tuite,M.F. (1997). The non-standard genetic code of Candida spp.: an evolving genetic code or a novel mechanism for adaptation? Mol. Microbiol. *26*, 423-431.

Scannell,D.R., Zill,O.A., Rokas,A., Payen,C., Dunham,M.J., Eisen,M.B., Rine,J., Johnston,M., and Hittinger,C.T. (2011). The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. G3. (Bethesda.) *1*, 11-25.

Scherer,S. and Magee,P.T. (1990). Genetics of Candida albicans. Microbiol. Rev. *54*, 226-241.

Schieltz,D.M. (1999). Preparing 2-D protein extracts from yeast. Methods Mol. Biol. *112*, 31-34.

Schimmel,P. (2008a). An editing activity that prevents mistranslation and connection to disease. J. Biol. Chem. *283*, 28777-28782.

Schimmel,P. (2008b). Development of tRNA synthetases and connection to genetic code and disease. Protein Sci. *17*, 1643-1652.

Schimmel,P. (2011). Mistranslation and its control by tRNA synthetases. Philos. Trans. R. Soc. Lond B Biol. Sci. *366*, 2965-2971.

Schimmel,P. and Ribas de,P.L. (1995). Transfer RNA: from minihelix to genetic code. Cell *81*, 983-986.

Schmeing,T.M. and Ramakrishnan,V. (2009). What recent ribosome structures have revealed about the mechanism of translation. Nature *461*, 1234-1242.

Schmitt,M.E., Brown,T.A., and Trumpower,B.L. (1990). A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. Nucleic Acids Res. *18*, 3091-3092.

Schneider-Poetsch,T., Usui,T., Kaida,D., and Yoshida,M. (2010). Garbled messages and corrupted translations. Nat. Chem. Biol. *6*, 189-198.

Schroder,M. and Kaufman,R.J. (2005). ER stress and the unfolded protein response. Mutat. Res. *569*, 29-63.

Schultz,D.W. and Yarus,M. (1994). Transfer RNA mutation and the malleability of the genetic code. J. Mol. Biol. *235*, 1377-1380.

Schultz,D.W. and Yarus,M. (1996). On malleability in the genetic code. J. Mol. Evol. *42*, 597-601.

Schwanhausser,B., Busse,D., Li,N., Dittmar,G., Schuchhardt,J., Wolf,J., Chen,W., and Selbach,M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337-342.

Selmecki,A., Forche,A., and Berman,J. (2006). Aneuploidy and isochromosome formation in drug-resistant Candida albicans. Science *313*, 367-370.

Selmecki,A., Forche,A., and Berman,J. (2010). Genomic plasticity of the human fungal pathogen Candida albicans. Eukaryot. Cell *9*, 991-1008.

Selmecki,A., Gerami-Nejad,M., Paulson,C., Forche,A., and Berman,J. (2008). An isochromosome confers drug resistance in vivo by amplification of two genes, ERG11 and TAC1. Mol. Microbiol. *68*, 624-641.

Selmecki,A.M., Dulmage,K., Cowen,L.E., Anderson,J.B., and Berman,J. (2009). Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. PLoS. Genet. *5*, e1000705.

Sengupta,S. and Higgs,P.G. (2005). A unified model of codon reassignment in alternative genetic codes. Genetics *170*, 831-840.

Sengupta,S., Yang,X., and Higgs,P.G. (2007). The mechanisms of codon reassignments in mitochondrial genetic codes. J. Mol. Evol. *64*, 662-688.

Shaul,S., Berel,D., Benjamini,Y., and Graur,D. (2010). Revisiting the operational RNA code for amino acids: Ensemble attributes and their implications. RNA. *16*, 141-153.

Shendure,J. and Ji,H. (2008). Next-generation DNA sequencing. Nat. Biotechnol. *26*, 1135-1145.

Siddik,Z.H. (2003). Cisplatin: mode of cytotoxic action and molecular basis of resistance. Oncogene *22*, 7265-7279.

Silva,R.M., Miranda,I., Moura,G., and Santos,M.A. (2004). Yeast as a model organism for studying the evolution of non-standard genetic codes. Brief. Funct. Genomic. Proteomic. *3*, 35-46.

Silva,R.M. *et al.* (2007). Critical roles for a genetic code alteration in the evolution of the genus Candida. EMBO J. *26*, 4555-4565.

Simon,R., Lam,A., Li,M.C., Ngan,M., Menenzes,S., and Zhao,Y. (2007). Analysis of gene expression data using BRB-ArrayTools. Cancer Inform. *3*, 11-17.

Singh,R.P., Prasad,H.K., Sinha,I., Agarwal,N., and Natarajan,K. (2011). Cap2-HAP complex is a critical transcriptional regulator that has dual but contrasting roles in regulation of iron homeostasis in Candida albicans. J. Biol. Chem. *286*, 25154-25170.

Singh,S.D., Robbins,N., Zaas,A.K., Schell,W.A., Perfect,J.R., and Cowen,L.E. (2009). Hsp90 governs echinocandin resistance in the pathogenic yeast Candida albicans via calcineurin
8. PLoS. Pathog. *5*, e1000532.

Skrzypek,M.S., Arnaud,M.B., Costanzo,M.C., Inglis,D.O., Shah,P., Binkley,G., Miyasato,S.R., and Sherlock,G. (2010). New tools at the Candida Genome Database: biochemical pathways and full-text literature search. Nucleic Acids Res. *38*, D428-D432.

Slupska,M.M., Baikalov,C., Lloyd,R., and Miller,J.H. (1996). Mutator tRNAs are encoded by the Escherichia coli mutator genes mutA and mutC: a novel pathway for mutagenesis. Proc. Natl. Acad. Sci. U. S. A *93*, 4380-4385.

Slupska,M.M., King,A.G., Lu,L.I., Lin,R.H., Mao,E.F., Lackey,C.A., Chiang,J.H., Baikalov,C., and Miller,J.H. (1998). Examination of the role of DNA polymerase proofreading in the mutator effect of miscoding tRNAs. J. Bacteriol. *180*, 5712-5717.

Slutsky,B., Staebell,M., Anderson,J., Risen,L., Pfaller,M., and Soll,D.R. (1987). "White-opaque transition": a second high-frequency switching system in Candida albicans. J. Bacteriol. *169*, 189-197.

Smith,D.A., Nicholls,S., Morgan,B.A., Brown,A.J., and Quinn,J. (2004). A conserved stress-activated protein kinase regulates a core stress response in the human pathogen Candida albicans. Mol. Biol. Cell *15*, 4179-4190.

Smith,D.R. *et al.* (2008). Rapid whole-genome mutational profiling using next-generation sequencing technologies. Genome Res. *18*, 1638-1642.

Soll,D.R. (1992). High-frequency switching in Candida albicans. Clin. Microbiol. Rev. *5*, 183-203.

Soma,A. and Himeno,H. (1998). Cross-species aminoacylation of tRNA with a long variable arm between Escherichia coli and Saccharomyces cerevisiae. Nucleic Acids Res. *26*, 4374-4381.

Soma,A., Kumagai,R., Nishikawa,K., and Himeno,H. (1996). The anticodon loop is a major identity determinant of Saccharomyces cerevisiae tRNA(Leu). J. Mol. Biol. *263*, 707-714.

Sorgo,A.G., Heilmann,C.J., Dekker,H.L., Bekker,M., Brul,S., de Koster,C.G., de Koning,L.J., and Klis,F.M. (2011). Effects of fluconazole on the secretome, the wall proteome, and wall integrity of the clinical fungus Candida albicans. Eukaryot. Cell *10*, 1071-1081.

Sprinzl,M. and Cramer,F. (1975). Site of aminoacylation of tRNAs from Escherichia coli with respect to the 2'- or 3'-hydroxyl group of the terminal adenosine. Proc. Natl. Acad. Sci. U. S. A *72*, 3049-3053.

Sprinzl,M., Horn,C., Brown,M., Ioudovitch,A., and Steinberg,S. (1998). Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. *26*, 148-153.

Srinivasan,G., James,C.M., and Krzycki,J.A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science *296*, 1459-1462.

Srivatsan,A., Han,Y., Peng,J., Tehranchi,A.K., Gibbs,R., Wang,J.D., and Chen,R. (2008). High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. PLoS. Genet. *4*, e1000139.

Stansfield,I., Jones,K.M., Herbert,P., Lewendon,A., Shaw,W.V., and Tuite,M.F. (1998). Missense translation errors in Saccharomyces cerevisiae. J. Mol. Biol. *282*, 13-24.

Stathopoulos,C., Li,T., Longman,R., Vothknecht,U.C., Becker,H.D., Ibba,M., and Soll,D. (2000). One polypeptide with two aminoacyl-tRNA synthetase activities. Science *287*, 479-482.

Stoltzfus,A. and Yampolsky,L.Y. (2007). Amino acid exchangeability and the adaptive code hypothesis. J. Mol. Evol. *65*, 456-462.

Su,D., Lieberman,A., Lang,B.F., Simonovic,M., Soll,D., and Ling,J. (2011). An unusual tRNAThr derived from tRNAHis reassigns in yeast mitochondria the CUN codons to threonine. Nucleic Acids Res. *39*, 4866-4874.

Sudbery,P., Gow,N., and Berman,J. (2004). The distinct morphogenic states of Candida albicans. Trends Microbiol. *12*, 317-324.

Sugita,T. and Nakase,T. (1999). Nonuniversal usage of the leucine CUG codon in yeasts: Investigation of basidiomycetous yeast. J. Gen. Appl. Microbiol. *45*, 193-197.

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. *16*, 1122-1134.

Suzuki,T., Ueda,T., Yokogawa,T., Nishikawa,K., and Watanabe,K. (1994). Characterization of serine and leucine tRNAs in an asporogenic yeast Candida cylindracea and evolutionary implications of genes for tRNA(Ser)CAG responsible for translation of a non-universal genetic code. Nucleic Acids Res. *22*, 115-123.

Taylor,F.J. and Coates,D. (1989). The code within the codons. Biosystems *22*, 177-187.

Teo,S.M., Pawitan,Y., Ku,C.S., Chia,K.S., and Salim,A. (2012). Statistical challenges associated with detecting copy number variations with next-generation sequencing. Bioinformatics.

Tosa,T. and Pizer,L.I. (1971). Biochemical bases for the antimetabolite action of L-serine hydroxamate. J. Bacteriol. *106*, 972-982.

Trifonov,E.N. (1989). The multiple codes of nucleotide sequences. Bull. Math. Biol. *51*, 417-432.

Trotter,E.W., Kao,C.M., Berenfeld,L., Botstein,D., Petsko,G.A., and Gray,J.V. (2002). Misfolded proteins are competent to mediate a subset of the responses to heat shock in Saccharomyces cerevisiae. J. Biol. Chem. *277*, 44817-44825.

Tsai,I.J., Bensasson,D., Burt,A., and Koufopanou,V. (2008). Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle. Proc. Natl. Acad. Sci. U. S. A *105*, 4957-4962.

Tsong,A.E., Tuch,B.B., Li,H., and Johnson,A.D. (2006). Evolution of alternative transcriptional circuits with identical logic. Nature *443*, 415-420.

Tuite,M.F. and Santos,M.A. (1996). Codon reassignment in Candida species: an evolutionary conundrum. Biochimie *78*, 993-999.

Turanov,A.A., Lobanov,A.V., Fomenko,D.E., Morrison,H.G., Sogin,M.L., Klobutcher,L.A., Hatfield,D.L., and Gladyshev,V.N. (2009). Genetic code supports targeted insertion of two amino acids by one codon. Science *323*, 259-261.

van der Gulik,P.T. and Hoff,W.D. (2011). Unassigned codons, nonsense suppression, and anticodon modifications in the evolution of the genetic code. J. Mol. Evol. *73*, 59-69.

Venkatraman,E.S. and Olshen,A.B. (2007). A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics. *23*, 657-663.

Vogel,C., Abreu,R.S., Ko,D., Le,S.Y., Shapiro,B.A., Burns,S.C., Sandhu,D., Boutz,D.R., Marcotte,E.M., and Penalva,L.O. (2010). Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. Mol. Syst. Biol. *6*, 400.

Vogel,C. and Marcotte,E.M. (2008). Calculating absolute and relative protein abundance from mass spectrometry-based protein expression data. Nat. Protoc. *3*, 1444-1451.

Vogel,C. and Marcotte,E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat. Rev. Genet. *13*, 227-232.

Vogel,C., Silva,G.M., and Marcotte,E.M. (2011). Protein expression regulation under oxidative stress. Mol. Cell Proteomics. *10*, M111.

Walther,A. and Wendland,J. (2003). An improved transformation protocol for the human fungal pathogen Candida albicans. Curr. Genet. *42*, 339-343.

Wang,L. and Schultz,P.G. (2001). A general approach for the generation of orthogonal tRNAs. Chem. Biol. *8*, 883-890.

Wang,L., Xie,J., and Schultz,P.G. (2006a). Expanding the genetic code. Annu. Rev. Biophys. Biomol. Struct. *35*, 225-249.

Wang,Q., Parrish,A.R., and Wang,L. (2009). Expanding the genetic code for biological studies. Chem. Biol. *16*, 323-336.

Wang,Y., Cao,Y.Y., Jia,X.M., Cao,Y.B., Gao,P.H., Fu,X.P., Ying,K., Chen,W.S., and Jiang,Y.Y. (2006b). Cap1p is involved in multiple pathways of oxidative stress response in Candida albicans. Free Radic. Biol. Med. *40*, 1201-1209.

Watamoto,T., Samaranayake,L.P., Egusa,H., Yatani,H., and Seneviratne,C.J. (2011). Transcriptional regulation of drug-resistance genes in Candida albicans biofilms in response to antifungals. J. Med. Microbiol. *60*, 1241-1247.

Watanabe,Y., Tsurui,H., Ueda,T., Furushima,R., Takamiya,S., Kita,K., Nishikawa,K., and Watanabe,K. (1994). Primary and higher order structures of nematode (Ascaris suum) mitochondrial tRNAs lacking either the T or D stem. J. Biol. Chem. *269*, 22902-22906.

Weaver,R.F. (2001). Molecular Biology, Lawrence: The McGraw-Hill Companies.

Webster,M., Witkin,K.L., and Cohen-Fix,O. (2009). Sizing up the nucleus: nuclear shape, size and nuclear-envelope assembly. J. Cell Sci. *122*, 1477-1486.

Weissman,Z., Berdicevsky,I., Cavari,B.Z., and Kornitzer,D. (2000). The high copper tolerance of Candida albicans is mediated by a P-type ATPase
1. Proc. Natl. Acad. Sci. U. S. A *97*, 3520-3525.

Wellington,M., Kabir,M.A., and Rustchenko,E. (2006). 5-fluoro-orotic acid induces chromosome alterations in genetically manipulated strains of Candida albicans. Mycologia. *98*, 393-398.

Wilson,D.N. and Nierhaus,K.H. (2003). The ribosome through the looking glass. Angew. Chem. Int. Ed Engl. *42*, 3464-3486.

Woese,C.R. (1965). On the evolution of the genetic code. Proc. Natl. Acad. Sci. U. S. A *54*, 1546-1552.

Woese,C.R., Olsen,G.J., Ibba,M., and Soll,D. (2000). Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiol. Mol. Biol. Rev. *64*, 202-236.

Wong,J.T. (1975). A co-evolution theory of the genetic code. Proc. Natl. Acad. Sci. U. S. A *72*, 1909-1912.

Wong,J.T. and Bronskill,P.M. (1979). Inadequacy of prebiotic synthesis as origin of proteinous amino acids. J. Mol. Evol. *13*, 115-125.

Xiang,Q. and Glass,N.L. (2004). Chromosome rearrangements in isolates that escape from het-c heterokaryon incompatibility in Neurospora crassa. Curr. Genet. *44*, 329-338.

Xu,D. *et al.* (2007). Genome-wide fitness test and mechanism-of-action studies of inhibitory compounds in Candida albicans
10. PLoS. Pathog. *3*, e92.

Yamao,F., Muto,A., Kawauchi,Y., Iwami,M., Iwagami,S., Azumi,Y., and Osawa,S. (1985). UGA is read as tryptophan in Mycoplasma capricolum. Proc. Natl. Acad. Sci. U. S. A *82*, 2306-2309.

Yarus,M., Caporaso,J.G., and Knight,R. (2005). Origins of the genetic code: the escaped triplet theory. Annu. Rev. Biochem. *74*, 179-198.

Yin,Z., Stead,D., Selway,L., Walker,J., Riba-Garcia,I., McLnerney,T., Gaskell,S., Oliver,S.G., Cash,P., and Brown,A.J. (2004). Proteomic response to amino acid starvation in Candida albicans and Saccharomyces cerevisiae. Proteomics. *4*, 2425-2436.

Yin,Z., Stead,D., Walker,J., Selway,L., Smith,D.A., Brown,A.J., and Quinn,J. (2009). A proteomic analysis of the salt, cadmium and peroxide stress responses in Candida albicans and the role of the Hog1 stress-activated MAPK in regulating the stress-induced proteome. Proteomics. *9*, 4686-4703.

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. *53*, 314-326.

Yokoyama,S. and Nishimura,S. (1995). Modified Nucleosides and Codon Recognition. In: tRNA structure, Biosynthesis and function., ed. D.Soll and U.RajBhandaryWashington,D.C.: American Society for Microbiology, 207-223.

Young,T.S. and Schultz,P.G. (2010). Beyond the canonical 20 amino acids: expanding the genetic lexicon. J. Biol. Chem. *285*, 11039-11044.

Yuan,J., O'Donoghue,P., Ambrogelly,A., Gundllapalli,S., Sherrer,R.L., Palioura,S., Simonovic,M., and Soll,D. (2010). Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems. FEBS Lett. *584*, 342-349.

Zaher,H.S. and Green,R. (2009). Quality control by the ribosome following peptide bond formation. Nature *457*, 161-166.

Zerbino,D.R. and Birney,E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. *18*, 821-829.

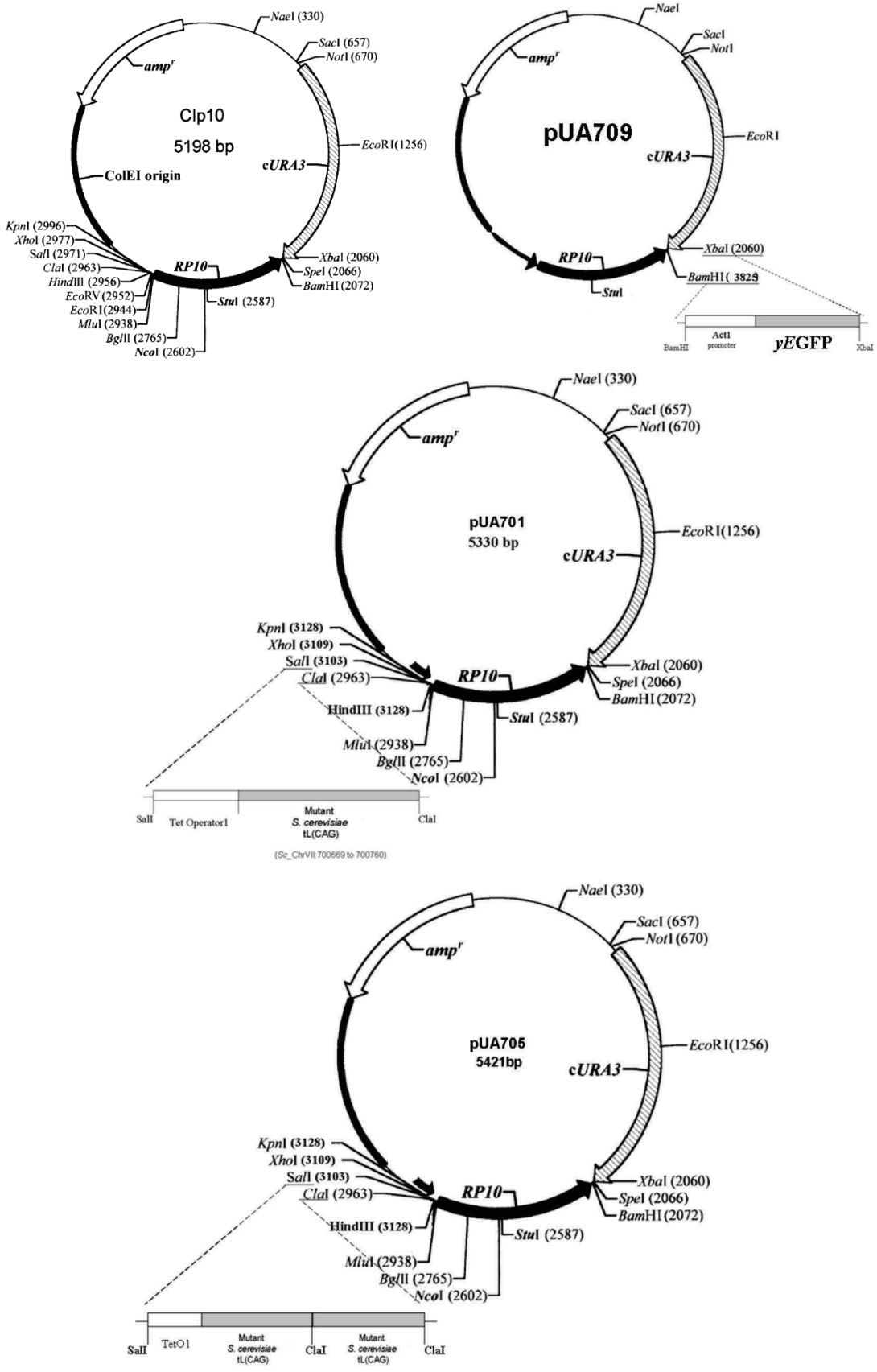
Zhao,J., Leung,H.E., and Winkler,M.E. (2001). The miaA mutator phenotype of Escherichia coli K-12 requires recombination functions. J. Bacteriol. *183*, 1796-1800.

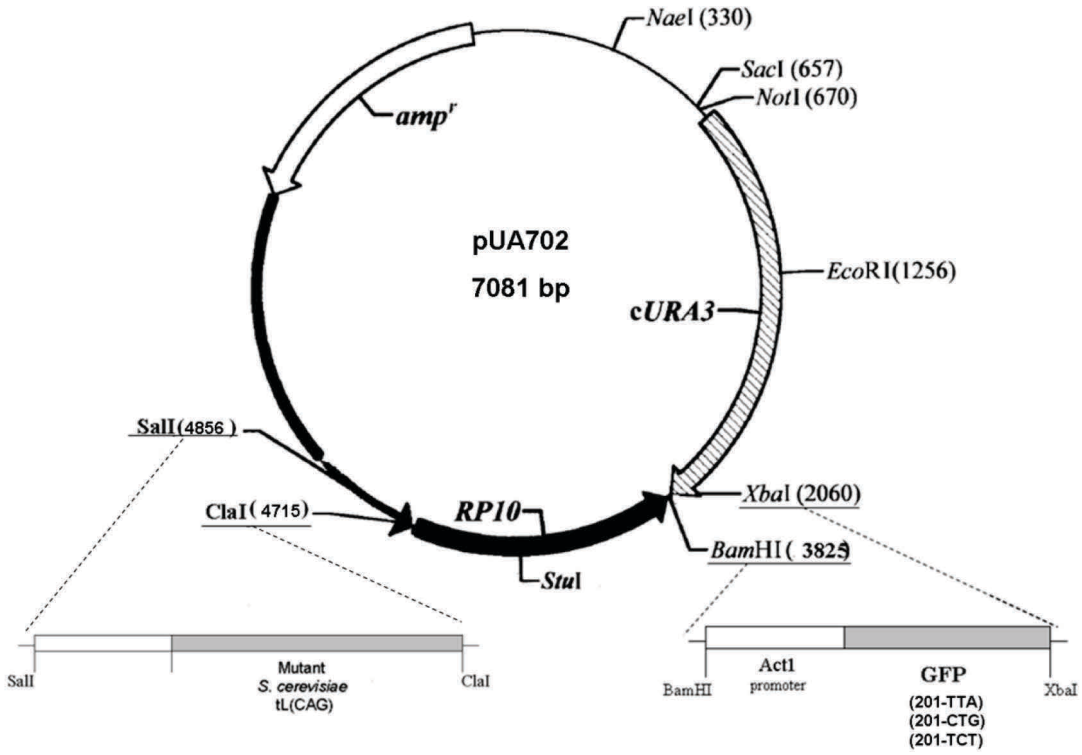
Znaidi,S., Barker,K.S., Weber,S., Alarco,A.M., Liu,T.T., Boucher,G., Rogers,P.D., and Raymond,M. (2009). Identification of the Candida albicans Cap1p regulon. Eukaryot. Cell *8*, 806-820.

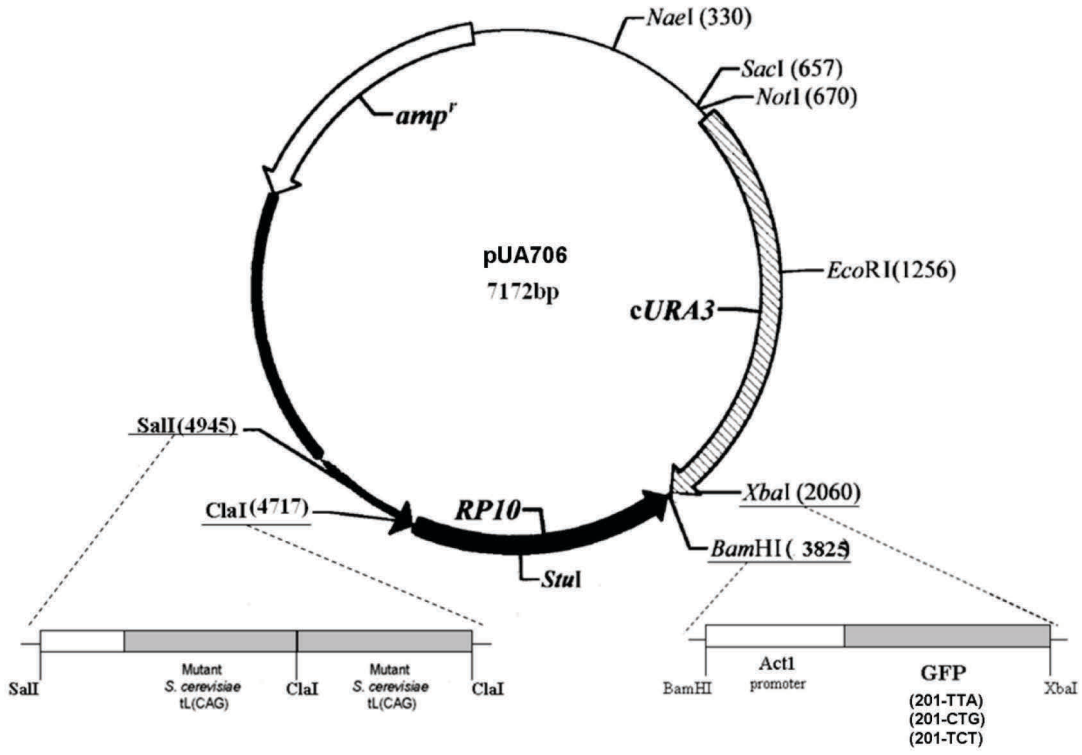
Zolan,M.E. (1995). Chromosome-length polymorphism in fungi. Microbiol. Rev. *59*, 686-698.

7. Annexes

Annex A: Map of the Plasmids

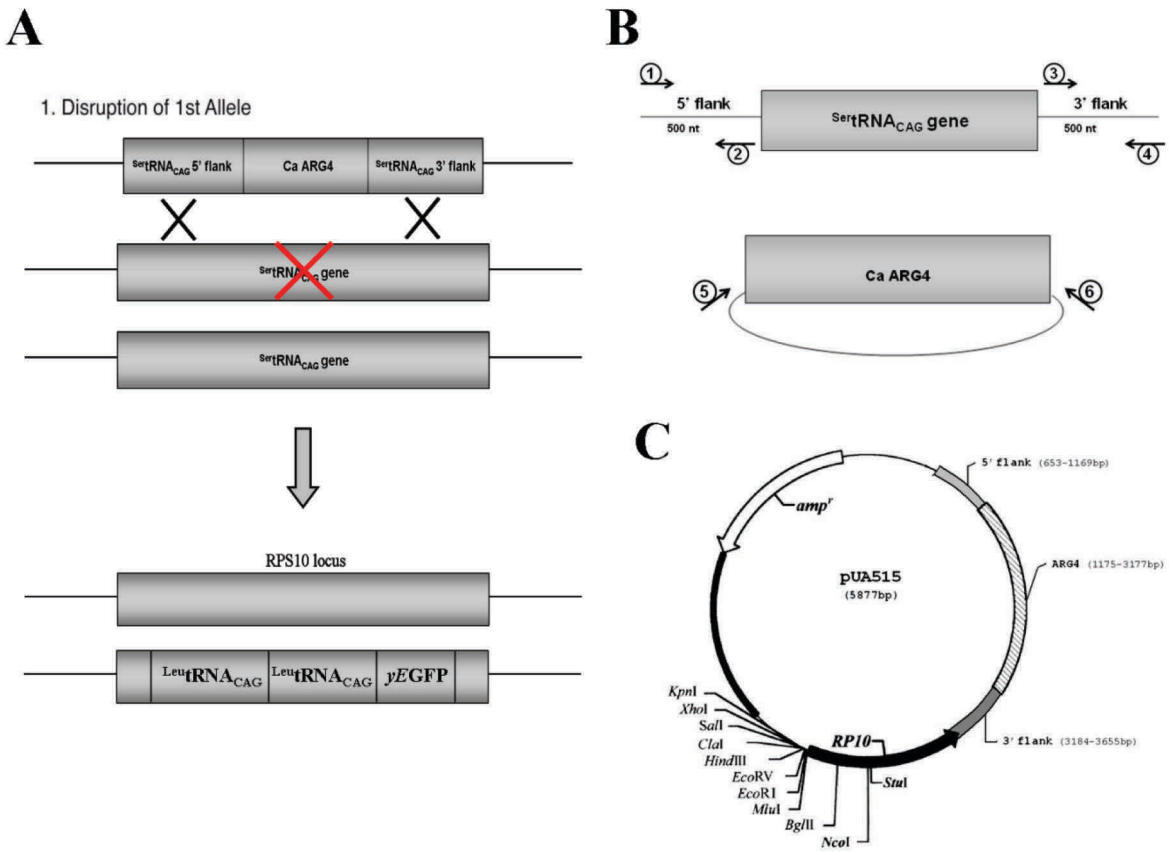



NaeI (330)
SacI (657)
NotI (670)
amp^r
pUA702
7081 bp
EcoRI(1256)
cURA3
SalI(4856)
ClaI(4715)
RP10
XbaI (2060)
BamHI(3825)
StuI
SalI
Mutant
S. cerevisiae
tL(CAG)
ClaI
BamHI
Act1
promoter
GFP
(201-TTA)
(201-CTG)
(201-TCT)
XbaI


NaeI (330)
SacI (657)
NotI (670)
amp^r
pUA706
7172bp
EcoRI(1256)
cURA3
SalI(4945)
ClaI(4717)
RP10
XbaI (2060)
BamHI(3825)
StuI
SalI
Mutant
S. cerevisiae
tL(CAG)
ClaI
Mutant
S. cerevisiae
tL(CAG)
ClaI
BamHI
Act1
promoter
GFP
(201-TTA)
(201-CTG)
(201-TCT)
XbaI

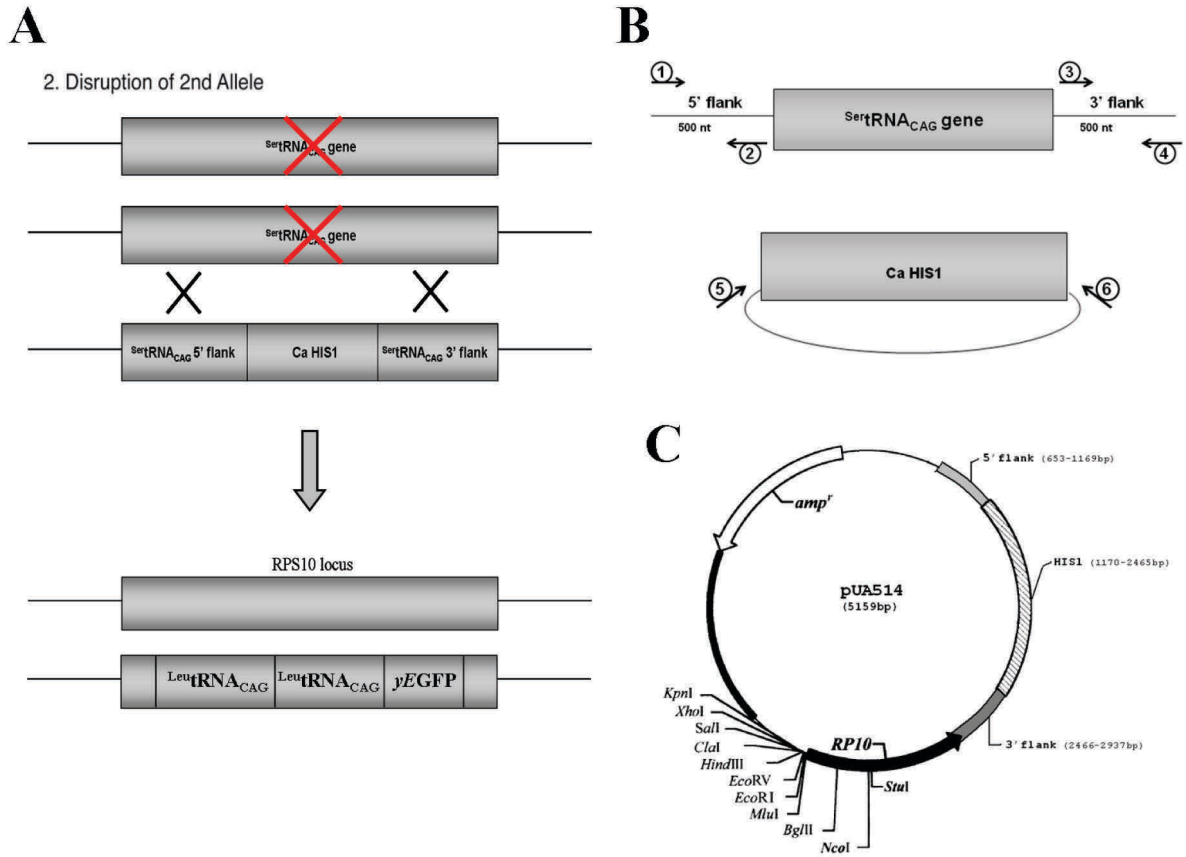
Annexe B: Disruption of the first allele of the tRNA$^{Ser}_{CAG}$ gene

Strategy used to disrupt one copy of the tRNA$_{CAG}^{Ser}$ from T2 strain, producing the T2KO1 strain. **A.** Deletion was carried out using an ARG4 marker cassette tagged with a 500nt-flanking homology for homologous recombination. **B.** The tRNA$^{Ser}_{CAG}$ 5' and 3' flanking sequences were amplified by PCR with primers 1, 2 and 3,4, respectively. Primers 5, 6 were used to amplify the auxotrophic marker. **C.** The flanking sequences and the selectable marker were cloned into an empty plasmid (pUA515) that allowed for the complete assembly of the disruption fragment. Primers 1 and 4 were used to to amplify the *tScag::ARG4* fragment by standard PCR protocols. Strain T2 was transformed with the *tScag::ARG4* cassette originating strains T2KO1.



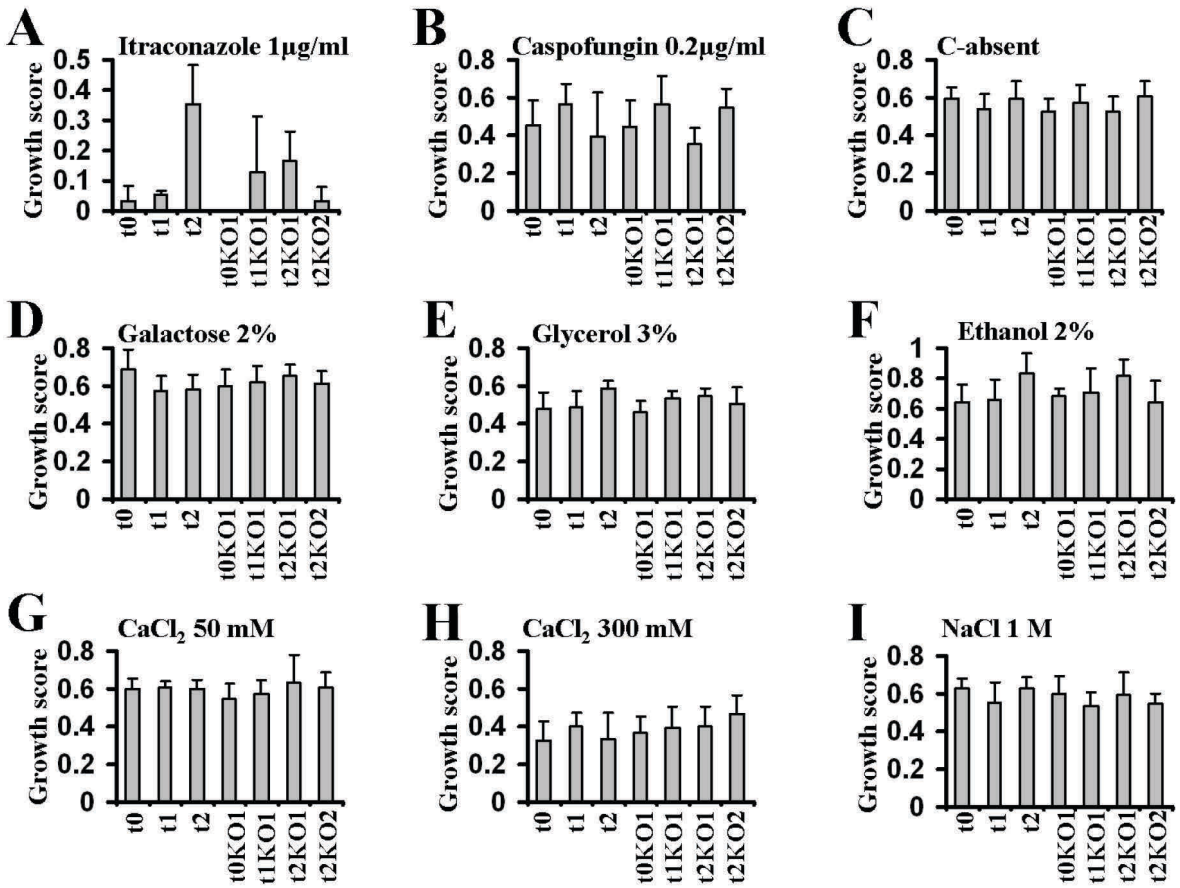
Annexe C: Disruption of the second allele of the tRNA$^{Ser}_{CAG}$ gene

Strategy used to disrupt the second copy of the tRNA$_{CAG}^{Ser}$ from T2KO1 strain, producing the T2KO2 strain. **A.** Deletion was made using an HIS1 marker cassette tagged with 500nt-flanking homology for homologous recombination. **B.** The tRNA$^{Ser}_{CAG}$ 5' and 3' flanking sequences were amplified by PCR with primers 1, 2 and 3,4, respectively. Primers 5, 6 were used to amplify the auxotrophic marker. **C.** The flanking sequences and the selectable marker were cloned into an empty plasmid (pUA514) that allowed for the complete assembly of the disruption fragment. Primers 1 and 4 were used to amplify the *tScag::HIS1* fragment by standard PCR protocols. The heterozygote T2KO1 was transformed with *tScag::HIS1* to produce the null mutant T2KO2.



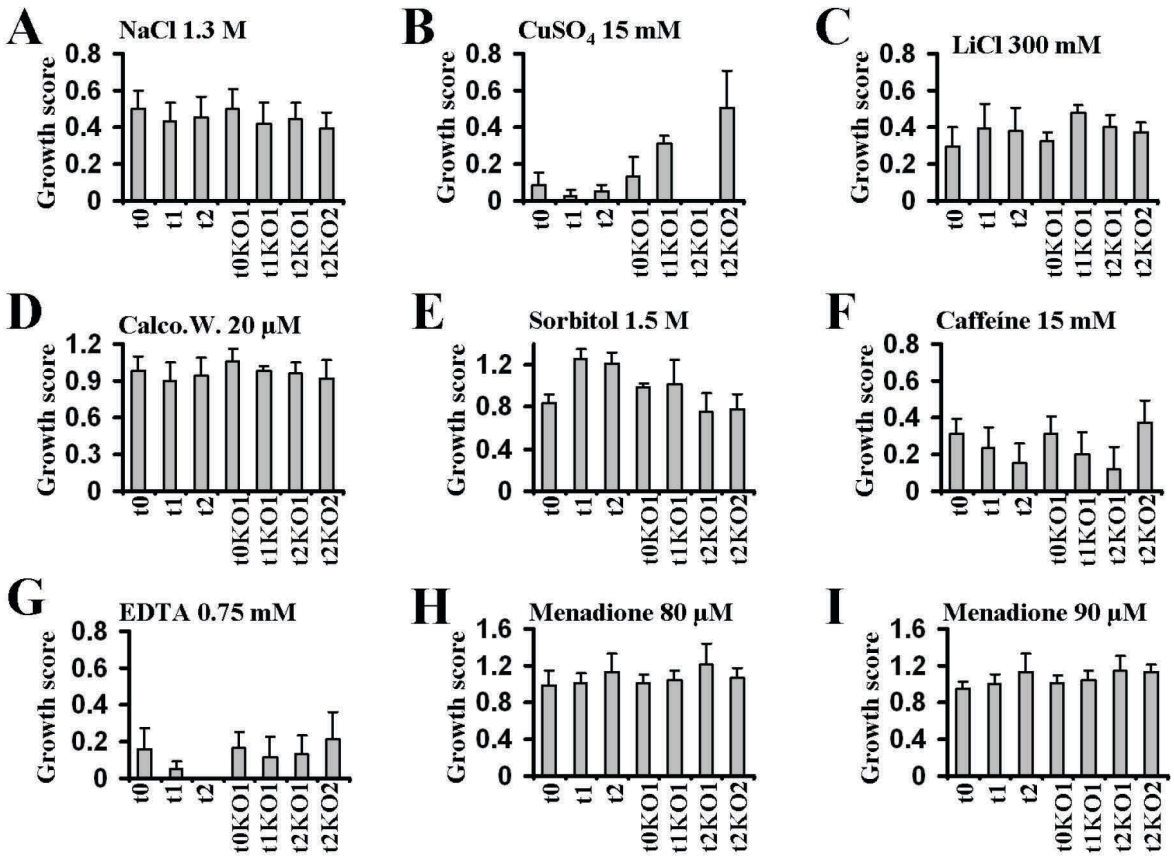
Annexe D: Growth scores of *C. albicans* strains under stress conditions

Growth scores represent the ratio between growth in the stress plate and the YPD/MM control plate. **A.** Growth scores in YPD medium supplemented with Itraconazole 1µg/ml. **B.** Caspofungin 0.2µg/ml. **C.** Growth scores in minimal medium without carbon source. **D.** 2% of galactose as carbon source. **E.** 3% glycerol. **F.** 2% ethanol. **G.** Growth scores in YPD medium supplemented with $CaCl_2$ 50 mM. **H.** $CaCl_2$ 50 mM. **I.** NaCl 1 M.



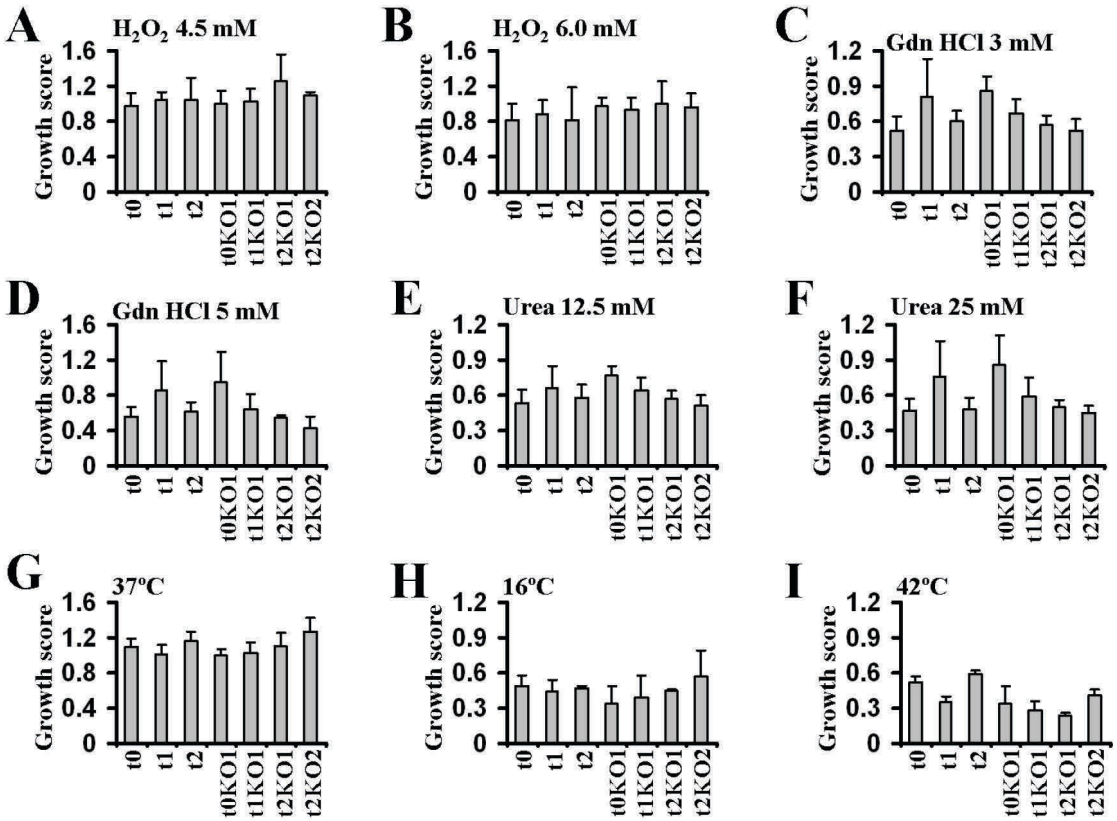
Annexe D: Growth scores of *C. albicans* strains under stress conditions

Growth scores represent the ratio between growth in the stress plate and the YPD/MM control plate. **A.** Growth scores in YPD medium supplemented with NaCl 1.3 M. **B.** $CuSO_4$ 15 mM. **C.** LiCl 300 mM. **D.** Calcofluor White 20 µM. **E.** Sorbitol 1.5 M. **F.** Caffeíne 15 mM. **G.** EDTA 0.75 mM. **H.** Menadione 80 µM. **I.** Menadione 90 µM.



Annexe D: Growth scores of *C. albicans* strains under stress conditions

Growth scores represent the ratio between growth in the stress plate and the YPD/MM control plate. **A.** Growth scores in YPD medium supplemented with H_2O_2 4.5 mM. **B.** H_2O_2 6.0 mM. **C.** Guanidine HCl 3 mM. **D.** Guanidine HCl 5 mM. **E.** Urea 12.5 mM. **F.** Urea 25 mM. **G.** Growth at 37ºC. **H.** Growth at 16ºC. **I.** Growth at 42ºC.



Annexe E: Phenotype MicroArray panels

Graphical depiction of yeast metabolic and sensitivity phenotype profiles. Figures represent an array of 96-well plate representations. **A.** PM1 and PM2 contain carbon utilization assays. **B.** PM3 contains nitrogen utilization assays. **C.** PM4 contains phosphorus and sulfur utilization assays. **D.** PM5 contains nutrient supplements for testing biosynthetic pathway/nutrient stimulation. **E.** PM6, PM7 and PM8 contain peptide nitrogen utilization assays. **F.** PM9 contains osmotic/ionic response assays. **G.** PM10 contains pH response assays. **H.** PM21to PM25 contain numerous yeast chemical sensitivity assays.

A. PM1-2 – Carbon Sources

	1	2	3	4	5	6	7	8	9	10	11	12
A	Negative control	L-Arabinose	N-Acetyl-D-Glucosamine	D-Saccharic Acid	Succinic Acid	D-Galactose	L-Aspartic Acid	L-Proline	D-Alanine	D-Trehalose	D-Mannose	Dulcitol
B	D-Serine	D-Sorbitol	Glycerol	L-Fucose	D-Glucuronic Acid	D-Gluconic Acid	D, L-α-Glycerol-Phosphate	D-Xylose	L-Lactic Acid	Formic Acid	D-Mannitol	L-Glutamic Acid
C	D-Glucose-6-Phospate	D-Galactonic Acid-γ-Lactone	D,L-Malic Acid	D-Ribose	Tween 20	L-Rhamnose	D-Fructose	Acetic Acid	α-D-Glucose	Maltose	D-Melibiose	Thymidine
D	L-Asparagine	D-Aspartic Acid	D-Glucosaminic Acid	1,2-Propanediol	Tween 40	α-Keto-Glutaric Acid	α-Keto-Butyric Acid	α-Methyl-D-Galactoside	α-D-Lactose	Lactulose	Sucrose	Uridine
E	L-Glutamine	M-Tartaric Acid	D-Glucose-1-Phosphate	D-Fructose-6-Phosphate	Tween 80	α-Hydroxy Glutaric Acid-γ-Lactone	α-Hydroxy Butyric Acid	β-Methyl-D-Glucoside	Adonitol	Maltotriose	2-Deoxy Adenosine	Adenosine
F	Glycyl-L-Aspartic Acid	Citric Acid	M-Inositol	D-Threonine	Fumaric Acid	Bromo Succinic Acid	Propionic Acid	Mucic Acid	Glycolic Acid	Glyoxylic Acid	D-Cellobiose	Inosine
G	Glycyl-L-Glutamic Acid	Tricarballylic Acid	L-Serine	L-Threonine	L-Alanine	L-Alanyl-Glycine	Acetoacetic Acid	N-Acetyl-β-D-Mannosamine	Mono Methyl Succinate	Methyl Pyruvate	D-Malic Acid	L-Malic Acid
H	Glycyl-L-Proline	p-Hydroxy Phenyl Acetic Acid	m-Hydroxy Phenyl Acetic Acid	Tyramine	D-Psicose	L-Lyxose	Glucuronamide	Pyruvic Acid	L-Galactonic Acid-γ-Lactone	D-Galacturonic Acid	Phenylethyl-amine	2-Aminoethanol

	1	2	3	4	5	6	7	8	9	10	11	12
A	Negative control	Chondroitin Sulfate C	α-Cyclodextrin	β-Cyclodextrin	γ-Cyclodextrin	Dextrin	Gelatin	Glycogen	Inulin	Laminarin	Mannan	Pectin
B	N-Acetyl-D-Galactosamine	N-Acetyl-Neuraminic Acid	β-D-Allose	Amygdalin	D-Arabinose	D-Arabitol	L-Arabitol	Arbutin	2-Deoxy-D-Ribose	I-Erythritol	D-Fucose	3-0-β-D-Galacto-pyranosyl-D-Arabinose
C	Gentiobiose	L-Glucose	Lactitol	D-Melezitose	Maltitol	α-Methyl-D-Glucoside	β-Methyl-D-Galactoside	3-Methyl Glucose	β-Methyl-D-Glucuronic Acid	α-Methyl-D-Mannoside	β-Methyl-D-Xyloside	Palatinose
D	D-Raffinose	Salicin	Sedoheptulosan	L-Sorbose	Stachyose	D-Tagatose	Turanose	Xylitol	N-Acetyl-D-Glucosaminitol	γ-Amino Butyric Acid	δ-Amino Valeric Acid	Butyric Acid
E	Capric Acid	Caproic Acid	Citraconic Acid	Citramalic Acid	D-Glucosamine	2-Hydroxy Benzoic Acid	4-Hydroxy Benzoic Acid	β-Hydroxy Butyric Acid	γ-Hydroxy Butyric Acid	α-Keto Valeric Acid	Itaconic Acid	5-Keto-D-Gluconic Acid
F	D-Lactic Acid	Methyl Ester	Malonic Acid	Melibionic Acid	Oxalic Acid	Oxalomalic Acid	Quinic Acid	D-Ribono-1,4-Lactone	Sebacic Acid	Sorbic Acid	Succinamic Acid	D-Tartaric Acid
G	Acetamide	L-Alaninamide	N-Acetyl-L-Glutamic Acid	L-Arginine	Glycine	L-Histidine	L-Homoserine	Hydroxy-L-Proline	L-Isoleucine	L-Leucine	L-Lysine	L-Methionine
H	L-Ornithine	L-Phenylalanine	L-Pyroglutamic Acid	L-Valine	D,L-Carnitine	Sec-Butylamine	D.L-Octopamine	Putrescine	Dihydroxy Acetone	2,3-Butanediol	2,3-Butanone	3-Hydroxy 2-Butanone

Annexe E: Phenotype MicroArray panels

B. PM3 – Nitrogen Sources

	1	2	3	4	5	6	7	8	9	10	11	12
A	Negative control	Ammonia	Nitrite	Nitrate	Urea	Biuret	L-Alanine	L-Arginine	L-Asparagine	L-Aspartic Acid	L-Cysteine	L-Glutamic Acid
B	L-Glutamine	Glycine	L-Histidine	L-Isoleucine	L-Leucine	L-Lysine	L-Methionine	L-Phenylalanine	L-Proline	L-Serine	L-Threonine	L-Tryptophan
C	L-Tyrosine	L-Valine	D-Alanine	D-Asparagine	D-Aspartic Acid	D-Glutamic Acid	D-Lysine	D-Serine	D-Valine	L-Citrulline	L-Homoserine	L-Ornithine
D	N-Acetyl-D,L-Glutamic Acid	N-Phthaloyl-L-Glutamic Acid	L-Pyroglutamic Acid	Hydroxylamine	Methylamine	N-Amylamine	N-Butylamine	Ethylamine	Ethanolamine	Ethylenediamine	Putrescine	Agmatine
E	Histamine	β-Phenylethyl-amine	Tyramine	Acetamide	Formamide	Glucuronamide	D,L-Lactamide	D-Glucosamine	D-Galactosamine	D-Mannosamine	N-Acetyl-D-Glucosamine	N-Acetyl-D-Galactosamine
F	N-Acetyl-D-Mannosamine	Adenine	Adenosine	Cytidine	Cytosine	Guanine	Guanosine	Thymine	Thymidine	Uracil	Uridine	Inosine
G	Xanthine	Xanthosine	Uric Acid	Alloxan	Allantoin	Parabanic Acid	D,L-α-Amino-N-Butyric Acid	γ-Amino-N-Butyric Acid	ε-Amino-N-Caproic Acid	D,L-α-Amino-Caprylic Acid	δ-Amino-N-Valeric Acid	α-Amino-N-Valeric Acid
H	Ala -Asp	Ala -Gln	Ala -Glu	Ala -Gly	Ala -His	Ala -Leu	Ala -Thr	Gly -Asn	Gly -Gln	Gly -Glu	Gly -Met	Met -Ala

C. PM4 – Phosphorus and Sulfur Sources

	1	2	3	4	5	6	7	8	9	10	11	12
A	Negative control	Phosphate	Pyrophosphate	Trimetaphosphate	Tripolyphosphate	Triethyl Phosphate	Hypophosphite	Adenosine-2'-monophosphate	Adenosine-3'-monophosphate	Adenosine-5'-monophosphate	Adenosine-2',3'-cyclic monophosphate	Adenosine-2',5'-cyclic monophosphate
B	Thiophosphate	Dithiophosphate	D,L-α-Glycerol Phosphate	β-Glycerol Phosphate	Carbamyl Phosphate	D-2-Phospho-Glyceric Acid	D-3-Phospho-Glyceric Acid	Guanosine-2'-monophosphate	Guanosine-3'-monophosphate	Guanosine-5'-monophosphate	Guanosine-2',3'-cyclic monophosphate	Guanosine -3',5'-cyclic monophosphate
C	Phosphoenol Pyruvate	Phospho-Glycolic Acid	D-Glucose-1-Phosphate	D-Glucose-6-Phosphate	2-Deoxy-D-Glucose 6-Phosphate	D-Glucosamine-6-Phosphate	6-Phospho-Gluconic Acid	Cytidine-2'-monophosphate	Cytidine-3'-monophosphate	Cytidine-5'-monophosphate	Cytidine-2',3'-cyclic monophosphate	Cytidine-2',5'-cyclic monophosphate
D	D-Mannose-1-Phosphate	D-Mannose-6-Phosphate	Cysteamine-S-Phosphate	Phospho-L-Arginine	O-Phospho-D-Serine	O-Phospho-L-Serine	O-Phospho-L-Threonine	Uridine-2' -monophosphate	Uridine-3' -monophosphate	Uridine-5' -monophosphate	Uridine-2',3'-cyclic monophosphate	Uridine-2',5'-cyclic monophosphate
E	O-Phospho-D-Tyrosine	O-Phospho-L-Tyrosine	Phosphocreatine	Phosphoryl Choline	O-Phosphoryl-Ethanolamine	Phosphono Acetic Acid	2-Aminoethyl Phosphonic Acid	Methylene Diphosphonic Acid	Thymidine-3'-monophosphate	Thymidine-5'-monophosphate	Inositol Hexaphosphate	Thymidine 3', 5'-cyclic monophosphate
F	Negative control	Sulfate	Thiosulfate	Tetrathionate	Thiophosphate	Dithiophosphate	L-Cysteine	D-Cysteine	L-Cysteinyl-Glycine	L-Cysteic Acid	Cysteamine	L-Cysteine Sulfinic Acid
G	N-Acetyl-L-Cysteine	S-Methyl-L-Cysteine	Cystathionine	Lanthionine	Glutathione	D,L-Ethionine	L-Methionine	D-Methionine	Glycyl-L-Methionine	N-Acetyl-D,L-Methionine	L-Methionine Sulfoxide	L-Methionine Sulfone
H	L-Djenkolic Acid	Thiourea	1-Thio-β-D-Glucose	D, L-Lipoamide	Taurocholic Acid	Taurine	Hypotaurine	p-Amino Benzene Sulfonic Acid	Butane Sulfonic Acid	2-Hydroxyethane Sulfonic Acid	Methane Sulfonic Acid	Tetramethylene Sulfone

D. PM5 – Nutrient Supplements

	1	2	3	4	5	6	7	8	9	10	11	12
A	Negative control	Positive Control	L-Alanine	L-Arginine	L-Asparagine	L-Aspartic Acid	L-Cysteine	L-Glutamic Acid	Adenosine-3',5'-cyclic monophosphate	Adenine	Adenosine	2'-Deoxy Adenosine
B	L-Glutamine	Glycine	L-Histidine	L-Isoleucine	L-Leucine	L-Lysine	L-Methionine	L-Phenylalanine	Guanosine -3',5'-cyclic monophosphate	Guanine	Guanosine	2'-Deoxy Guanosine
C	L-Proline	L-Serine	L-Threonine	L-Tryptophan	L-Tyrosine	L-Valine	L-isoleucine + L-Valine	trans-4-Hydroxy L-Proline	(5) 4-Amino-Imidazole-4(5)-Carboxamide	Hypoxanthine	Inosine	2'-Deoxy Inosine
D	L-Ornithine	L-Citrulline	Chorismic Acid	(-)Shikimic Acid	L-Homoserine Lactone	D-Alanine	D-Aspartic Acid	D-Glutamic Acid	D,L-α, ε-Diamino-pimelic Acid	Cytosine	Cytidine	2'-Deoxy Cytidine
E	Putrescine	Spermidine	Spermine	Pyridoxine	Pyridoxal	Pyridoxamine	β-Alanine	D-Pantothenic Acid	Orotic Acid	Uracil	Uridine	2'-Deoxy Uridine
F	Quinolinic Acid	Nicotinic Acid	Nicotinamide	β-Nicotinamide Adenine	δ-Amino-Levulinic Acid	Hematin	Deferoxamine Mesylate	D-(+)-Glucose	N-Acetyl D-Glucosamine	Thymine	Glutathione (reduced form)	Thymidine
G	Oxaloacetic Acid	D-Biotin	Cyano - Cobalamine	p-Amino-Benzoic Acid	Folic Acid	Inosine + Thiamine	Thiamine	Thiamine Pyrophosphate	Riboflavin	Pyrrolo-Quinoline Quinone	Menadione	Myo-Inositol
H	Butyric Acid	D,L-α-Hydroxy-Butyric Acid	α-Ketobutyric Acid	Caprylic Acid	D,L-α-Lipoic Acid (oxidized form)	D, L-Mevalonic Acid	D, L-Carnitine	Choline	Tween 20	Tween 40	Tween 60	Tween 80

Annexe E: Phenotype MicroArray panels

E. PM6-8 – Peptide Nitrogen Sources

	1	2	3	4	5	6	7	8	9	10	11	12
A	Negative control	Positive Control: L-Glutamine	Ala -Ala	Ala -Arg	Ala -Asn	Ala -Glu	Ala -Gly	Ala -His	Ala -Leu	Ala -Lys	Ala -Phe	Ala -Pro
B	Ala -Ser	Ala -Thr	Ala -Trp	Ala -Tyr	Arg-Ala	Arg-Arg	Arg-Asp	Arg-Gln	Arg-Glu	Arg-Ile	Arg-Leu	Arg-Lys
C	Arg-Met	Arg-Ph e	Arg-Se r	Arg-Trp	Arg-Ty r	Arg-Va l	Asn -Glu	Asn -V al	Asp -A sp	Asp -Glu	Asp -Le u	Asp -Lys
D	Asp -P he	Asp -T rp	Asp -V al	Cys -Gly	Gl n-G ln	Gl n-G ly	Gl u-Asp	Gl u-G lu	Gl u-G ly	Gl u-Se r	Gl u-Trp	Gl u-Ty r
E	Gl u-Va l	Gly -Ala	Gly -Arg	Gly -Cys	Gly -Gl y	Gly -His	Gly -L eu	Gly -L ys	Gly -Met	Gly -Phe	Gly -Pro	Gly -Se r
F	Gly -Th r	Gly -Trp	Gly -Tyr	Gly -Val	His -Asp	His -Gly	His -L eu	His -Ly s	His -Me t	His -Pro	His -Se r	His -T rp
G	His -T yr	His -Val	Ile -Ala	Ile -Arg	Ile -Gl n	Ile -Gly	Ile -His	Ile -I le	Ile -Me t	Ile -Phe	Ile -Pro	Ile -Se r
H	Ile -T rp	Ile -T yr	Ile -Val	Leu -A la	Leu -A rg	Leu -As p	Leu -Glu	Leu -Gly	Leu -I le	Leu -Le u	Leu -Me t	Leu -P he

	1	2	3	4	5	6	7	8	9	10	11	12
A	Negative control	Positive Control: L-Glutamine	Leu -S er	Leu -T rp	Leu -V al	Lys-Al a	Lys-A rg	Lys-G lu	Lys-Il e	Lys-Leu	Lys-Lys	Lys-Ph e
B	Lys-P ro	Lys-Se r	Lys-Th r	Lys-Trp	Lys-Ty r	Lys-Va l	Me t -A rg	Me t -As p	Me t -Gln	Me t -Glu	Me t -Gly	Me t -H is
C	Me t -I le	Me t -Le u	Me t -Lys	Me t -Me t	Me t -P he	Me t -P ro	Me t -T rp	Me t -V al	Phe -Ala	Phe -Gly	Phe -I le	Phe -Phe
D	Phe -Pro	Phe -Se r	Phe -T rp	Pro-Al a	Pro-Asp	Pro-G ln	Pro-G ly	Pro-Hy p	Pro-Leu	Pro-Ph e	Pro-P ro	Pro-Ty r
E	Se r-Al a	Se r-G ly	Se r-Hi s	Se r-Leu	Se r-Met	Se r-Ph e	Se r-P ro	Se r-Se r	Se r-Ty r	Se r-Va l	Thr-Ala	Thr-Arg
F	Thr-Gl u	Thr-Gl y	Thr-L eu	Thr-Met	Thr-Pro	Trp -Ala	Trp -Arg	Trp -Asp	Trp -Gl u	Trp -Gl y	Trp -L eu	Trp -L ys
G	Trp -Phe	Trp -Se r	Trp -Trp	Trp -Tyr	Tyr-Ala	Tyr-Gl n	Tyr-Gl u	Tyr-Gly	Tyr-His	Tyr-L eu	Tyr-Ly s	Tyr-Phe
H	Tyr-T rp	Tyr-T yr	Val -Arg	Val -Asn	Val -Asp	Val -Gly	Val -His	Val -Ile	Val -L eu	Val -T yr	Val -Val	Ɣ-G lu -Gly

	1	2	3	4	5	6	7	8	9	10	11	12
A	Negative control	Positive Control: L-Glutamine	Ala -Asp	Ala -Gl n	Ala -l le	Ala -Me t	Ala -Val	Asp -A la	Asp -Gln	Asp -Gly	Gl u-Al a	Gly -Asn
B	Gly -Asp	Gly -Ile	His -Ala	His -Gl u	His -His	Ile -Asn	Ile -L eu	Leu -As n	Leu -H is	Leu -P ro	Leu -Ty r	Lys-Asp
C	Lys-G ly	Lys-Met	Me t -Th r	Me t -Ty r	Phe -Asp	Phe -Gl u	Gl n-G lu	Phe -Me t	Phe -T yr	Phe -Val	Pro-A rg	Pro-Asn
D	Pro-G lu	Pro-Ile	Pro-Lys	Pro-Se r	Pro-Trp	Pro-Va l	Se r-Asn	Se r-Asp	Se r-G ln	Se r-G lu	Thr-Asp	Thr-Gl n
E	Thr-Phe	Thr-Se r	Trp -Val	Tyr-I le	Tyr-Val	Val -Ala	Val -Gl n	Val -Gl u	Val -Ly s	Val -Me t	Val -Phe	Val -Pro
F	Val -Se r	β-A la -A la	β-A la -Gly	β-A la -His	Me t -β-Al a	β-A la -P he	D-Al a-D -Ala	D-Al a-G ly	D-Al a-Leu	D-Leu -D -Le u	D-Leu -Gl y	D-Leu -Tyr
G	Ɣ-G lu -Gly	Ɣ-D -Glu -Gl y	Gly -D -A la	Gly -D -As p	Gly -D -S er	Gly -D -Th r	Gly -D -V al	Leu -β-Al a	Leu -D -L eu	Phe -β -A la	Ala -Ala -Ala	D-Al a-G ly -Gly
H	Gly -Gl y-Al a	Gly -Gl y-D -Le u	Gly -Gl y-G ly	Gly -Gl y-Ile	Gly -Gl y-Leu	Gly -Gl y-Ph e	Val -T yr-Val	Gly -Phe -Phe	Leu -Gly -Gly	Leu -Le u-Leu	Phe -Gly -Gl y	Tyr-Gly -Gl y

Annexe E: Phenotype MicroArray panels

F. PM9 – Osmolytes

	1	2	3	4	5	6	7	8	9	10	11	12
A	NaCl 1 %	NaCl 2 %	NaCl 3 %	NaCl 4 %	NaCl 5 %	NaCl 5 .5 %	NaCl 6 %	NaCl 6 .5 %	NaCl 7 %	NaCl 8 %	NaCl 9 %	NaCl 10 %
B	NaCl 6 %	NaCl 6 % + Betaine	NaCl 6 % + N-N Dimethyl glycine	NaCl 6 % + Sarcosine	NaCl 6 % + Dimethylsulphonyl propionate	NaCl 6 % + MOPS	NaCl 6 % + Ectoine	NaCl 6 % + Choline	NaCl 6% + Phosphoryl	NaCl 6 % + Creatine	NaCl 6 % + Creatinine	NaCl 6 % + L-Carnitine
C	NaCl 6 % + KCl	NaCl 6 % + L-p roline	NaCl 6 % + N-Acethyl L-glutamine	NaC1 6 % + β-Glutamic acid	NaCl 6 % + γ–Amino-n-butyric acid	NaC1 6 % + Glutathione	NaCl 6 % + Glycerol	NaC1 6 % + Trehalose	NaC1 6 % + Trimethylamine-N-oxide	NaC1 6 % + Trimethylamine	NaCl 6 % + Octopine	NaC1 6 % + Trigonelline
D	Potassium chloride 3%	Potassium chloride 4%	Potassium chloride 5%	Potassium chloride 6%	Sodium sulfate 2%	Sodium sulfate 3%	Sodium sulfate 4%	Sodium sulfate 5%	Ethylene glycol 5%	Ethylene glycol 10%	Ethylene glycol 15%	Ethylene glycol 20%
E	Sodium formate 1%	Sodium formate 2%	Sodium formate 3%	Sodium formate 4%	Sodium formate 5%	Sodium formate 6%	Urea 2%	Urea 3%	Urea 4%	Urea 5%	Urea 6%	Urea 7%
F	Sodium Lactate 1%	Sodium Lactate 2%	Sodium Lactate 3%	Sodium Lactate 4%	Sodium Lactate 5%	Sodium Lactate 6%	Sodium Lactate 7%	Sodium Lactate 8%	Sodium Lactate 9%	Sodium Lactate 10%	Sodium Lactate 11%	Sodium Lactate 12%
G	Sodium Phosphate pH7 20mM	Sodium Phosphate pH7 50mM	Sodium Phosphate pH7 100mM	Sodium Phosphate pH7 200mM	Sodium Benzoate pH5.2 20mM	Sodium Benzoate pH5.2 50mM	Sodium Benzoate pH5.2 100mM	Sodium Benzoate pH5.2 200mM	Ammonium sulfate pH8 10mM	Ammonium sulfate pH8 20mM	Ammonium sulfate pH8 50mM	Ammonium sulfate pH8 100mM
H	Sodium Nitrate 10mM	Sodium Nitrate 20mM	Sodium Nitrate 40mM	Sodium Nitrate 60mM	Sodium Nitrate 80mM	Sodium Nitrate 100mM	Sodium Nitrite 10mM	Sodium Nitrite 20mM	Sodium Nitrite 40mM	Sodium Nitrite 60mM	Sodium Nitrite 80mM	Sodium Nitrite 100mM

G. PM10 – pH

	1	2	3	4	5	6	7	8	9	10	11	12
A	pH 3. 5	pH 4	pH 4. 5	pH 5	pH 5. 5	pH 6	pH 7	pH 8	pH 8. 5	pH 9	pH 9. 5	pH 10
B	pH 4.5	pH 4.5 + L-Alanine	pH 4.5 + L-Arginine	pH 4.5 + L-Asparagine	pH 4.5 + L-Aspartic Acid	pH 4.5 + L-Glutamic Acid	pH 4.5 + L-Glutamine	pH 4.5 + Glycine	pH 4.5 + L-Histidine	pH 4.5 + L-Isoleucine	pH 4.5 + L-Leucine	pH 4.5 + L-Lysine
C	pH 4.5 + L-Methionine	pH 4.5 + L-Phenylalanine	pH 4.5 + L-Proline	pH 4.5 + L-Serine	pH 4.5 + L-Threonine	pH 4.5 + L-Tryptophan	pH 4.5 + L-Tyrosine	pH 4.5 + L-Valine	pH 4.5 + Hydroxy-L-Proline	pH 4.5 + L-Ornithine	pH 4.5 + L-Homoarginine	pH 4.5 + L-Homoserine
D	pH 4.5 + Anthranilic acid	pH 4.5 + L-Norleucine	pH 4.5 + L-Norvaline	pH 4.5 + α-Amino-N-butyric acid	pH 4.5 + p-Aminobenzoate	pH 4.5 + L-Cysteic acid	pH 4.5 + D-Lysine	pH 4.5 + 5-Hydroxy Lysine	pH 4.5 + 5-Hydroxy Tryptophan	pH 4.5 + D,L-Diamino pimelic acid	pH 4.5 + Trimethyl amine-N-oxide	pH 4.5 + Urea
E	pH 9.5	pH 9.5 + L-Alanine	pH 9.5 + L-Arginine	pH 9.5 + L-Asparagine	pH 9.5 + L-Aspartic Acid	pH 9.5 + L-Glutamic Acid	pH 9.5 + L-Glutamine	pH 9.5 + Glycine	pH 9.5 + L-Histidine	pH 9.5 + L-Isoleucine	pH 9.5 + L-Leucine	pH 9.5 + L-Lysine
F	pH 9.5 + L-Methionine	pH 9.5 + L-Phenylalanine	pH 9.5 + L-Proline	pH 9.5 + L-Serine	pH 9.5 + L-Threonine	pH 9.5 + L-Tryptophan	pH 9.5 + L-Tyrosine	pH 9.5 + L-Valine	pH 9.5 + Hydroxy-L-Proline	pH 9.5 + L-Ornithine	pH 9.5 + L-Homoarginine	pH 9.5 + L-Homoserine
G	pH 9.5 + Anthranilic acid	pH 9.5 + L-Norleucine	pH 9.5 + L-Norvaline	pH 9.5 + α-Amino-N-butyric acid	pH 9.5 + p-Aminobenzoate	pH 9.5 + L-Cysteic acid	pH 9.5 + D-Lysine	pH 9.5 + 5-Hydroxy Lysine	pH 9.5 + 5-Hydroxy Tryptophan	pH 9.5 + D,L-Diamino pimelic acid	pH 9.5 + Trimethyl amine-N-oxide	pH 9.5 + Urea
H	X-Caprylate	X–α-D-Glucoside	X-β-D-Glucoside	X-α -D-Galactoside	X-β -D - Galactoside	X-α-D-Glucuronide	X-β-D-Glucuronide	X-β-D-Glucosaminide	X-β-D-Galactosaminide	X-α-D-Mannoside	X-PO4	X-SO4

H. PM21 – Chemical Sensitivity

	1	2	3	4	5	6	7	8	9	10	11	12
A	Guanidine hydrochloride	Guanidine hydrochloride	Guanidine hydrochloride	Guanidine hydrochloride	2,2`-Dipyridyl	2,2`-Dipyridyl	2,2`-Dipyridyl	2,2`-Dipyridyl	Promethazine hydrochloride	Promethazine hydrochloride	Promethazine hydrochloride	Promethazine hydrochloride
B	Nystatin	Nystatin	Nystatin	Nystatin	Dodecyl trimethyl ammonium	Dodecyl trimethyl ammonium	Dodecyl trimethyl ammonium	Dodecyl trimethyl ammonium	Protamine Sulfate	Protamine Sulfate	Protamine Sulfate	Protamine Sulfate
C	Cetylpyridinium chloride	Cetylpyridinium chloride	Cetylpyridinium chloride	Cetylpyridinium chloride	Domiphen bromide	Domiphen bromide	Domiphen bromide	Domiphen bromide	L-Aspartic acid b-hydroxamate	L-Aspartic acid b-hydroxamate	L-Aspartic acid b-hydroxamate	L-Aspartic acid b-hydroxamate
D	1-Hydroxypyridine-2-thione	1-Hydroxypyridine-2-thione	1-Hydroxypyridine-2-thione	1-Hydroxypyridine-2-thione	EDTA	EDTA	EDTA	EDTA	Sodium Dichromate	Sodium Dichromate	Sodium Dichromate	Sodium Dichromate
E	Compound 48/80	Compound 48/81	Compound 48/82	Compound 48/83	Manganese(II) chloride	Manganese(II) chloride	Manganese(II) chloride	Manganese(II) chloride	Magnesium chloride	Magnesium chloride	Magnesium chloride	Magnesium chloride
F	Copper(II) Sulfate	Copper(II) Sulfate	Copper(II) Sulfate	Copper(II) Sulfate	Neomycin Sulfate	Neomycin Sulfate	Neomycin Sulfate	Neomycin Sulfate	D-Cycloserine	D-Cycloserine	D-Cycloserine	D-Cycloserine
G	Sodium Selenite	Sodium Selenite	Sodium Selenite	Sodium Selenite	Nickel chloride	Nickel chloride	Nickel chloride	Nickel chloride	Trifluoperazine	Trifluoperazine	Trifluoperazine	Trifluoperazine
H	Diamide	Diamide	Diamide	Diamide	Thiourea	Thiourea	Thiourea	Thiourea	Zinc chloride	Zinc chloride	Zinc chloride	Zinc chloride

Annexe E: Phenotype MicroArray panels

H. PM22-24 – Chemical Sensitivity

	1	2	3	4	5	6	7	8	9	10	11	12
A	L-Glutamic acid g hydroxamate	L-Glutamic acid g hydroxamate	L-Glutamic acid g hydroxamate	L-Glutamic acid g hydroxamate	Sodium Meta vanadate	Sodium Meta vanadate	Sodium Meta vanadate	Sodium Meta vanadate	Caffeine	Caffeine	Caffeine	Caffeine
B	L-Arginine hydroxamate	L-Arginine hydroxamate	L-Arginine hydroxamate	L-Arginine hydroxamate	Glycine hydroxamate	Glycine hydroxamate	Glycine hydroxamate	Glycine hydroxamate	Potassium Iodide	Potassium Iodide	Potassium Iodide	Potassium Iodide
C	3-Amino-1,2,4-Triazole	3-Amino-1,2,4-Triazole	3-Amino-1,2,4-Triazole	3-Amino-1,2,4-Triazole	Miltefosine	Miltefosine	Miltefosine	Miltefosine	D,L-Serine hydroxamate	D,L-Serine hydroxamate	D,L-Serine hydroxamate	D,L-Serine hydroxamate
D	Polymyxin B	Polymyxin B	Polymyxin B	Polymyxin B	Urea hydrogen peroxide	Urea hydrogen peroxide	Urea hydrogen peroxide	Urea hydrogen peroxide	Sodium Arsenate	Sodium Arsenate	Sodium Arsenate	Sodium Arsenate
E	Ceftriaxone	Ceftriaxone	Ceftriaxone	Ceftriaxone	BAPTA	BAPTA	BAPTA	BAPTA	D-Serine	D-Serine	D-Serine	D-Serine
F	Azaserine	Azaserine	Azaserine	Azaserine	Lithium chloride	Lithium chloride	Lithium chloride	Lithium chloride	Boric acid	Boric acid	Boric acid	Boric acid
G	Benzamidine	Benzamidine	Benzamidine	Benzamidine	Cycloheximide	Cycloheximide	Cycloheximide	Cycloheximide	Thallium(I) acetate	Thallium(I) acetate	Thallium(I) acetate	Thallium(I) acetate
H	Cephalothin	Cephalothin	Cephalothin	Cephalothin	Paromomycin	Paromomycin	Paromomycin	Paromomycin	Myclobutanil	Myclobutanil	Myclobutanil	Myclobutanil

	1	2	3	4	5	6	7	8	9	10	11	12
A	Benzethonium chloride	Benzethonium chloride	Benzethonium chloride	Benzethonium chloride	Chlorpromazine hydrochloride	Chlorpromazine hydrochloride	Chlorpromazine hydrochloride	Chlorpromazine hydrochloride	Ammonium Sulfate	Ammonium Sulfate	Ammonium Sulfate	Ammonium Sulfate
B	Cadmium chloride 2 1/2 H2O	Cadmium chloride 2 1/2 H2O	Cadmium chloride 2 1/2 H2O	Cadmium chloride 2 1/2 H2O	Dequalinium chloride	Dequalinium chloride	Dequalinium chloride	Dequalinium chloride	Doxycycline Hyclate	Doxycycline Hyclate	Doxycycline Hyclate	Doxycycline Hyclate
C	Glycine hydrochloride	Glycine hydrochloride	Glycine hydrochloride	Glycine hydrochloride	Hydroxylamine hydrochloride	Hydroxylamine hydrochloride	Hydroxylamine hydrochloride	Hydroxylamine hydrochloride	Poly-L-Lysine hydrochloride	Poly-L-Lysine hydrochloride	Poly-L-Lysine hydrochloride	Poly-L-Lysine hydrochloride
D	Chromium(III) chloride hexahydrate	Chromium(III) chloride hexahydrate	Chromium(III) chloride hexahydrate	Chromium(III) chloride hexahydrate	Cobalt(II) chloride hexahydrate	Cobalt(II) chloride hexahydrate	Cobalt(II) chloride hexahydrate	Cobalt(II) chloride hexahydrate	Cupric chloride dihydrate	Cupric chloride dihydrate	Cupric chloride dihydrate	Cupric chloride dihydrate
E	Sodium Meta borate hydrate	Sodium Meta borate hydrate	Sodium Meta borate hydrate	Sodium Meta borate hydrate	Sodium (meta) periodate	Sodium (meta) periodate	Sodium (meta) periodate	Sodium (meta) periodate	Sodium arsenite	Sodium arsenite	Sodium arsenite	Sodium arsenite
F	Sodium Azide	Sodium Azide	Sodium Azide	Sodium Azide	Sodium Caprylate	Sodium Caprylate	Sodium Caprylate	Sodium Caprylate	Sodium Cyanate	Sodium Cyanate	Sodium Cyanate	Sodium Cyanate
G	Sodium Nitrite	Sodium Nitrite	Sodium Nitrite	Sodium Nitrite	Sodium Ortho vanadate	Sodium Ortho vanadate	Sodium Ortho vanadate	Sodium Ortho vanadate	2-Deoxy-D-Glucose	2-Deoxy-D-Glucose	2-Deoxy-D-Glucose	2-Deoxy-D-Glucose
H	Sodium Selenate	Sodium Selenate	Sodium Selenate	Sodium Selenate	Sodium Cyanide	Sodium Cyanide	Sodium Cyanide	Sodium Cyanide	Sodium Thiosulfate	Sodium Thiosulfate	Sodium Thiosulfate	Sodium Thiosulfate

	1	2	3	4	5	6	7	8	9	10	11	12
A	Apramycin Sulfate	Apramycin Sulfate	Apramycin Sulfate	Apramycin Sulfate	9-Aminoacridine hydrochloride	9-Aminoacridine hydrochloride	9-Aminoacridine hydrochloride	9-Aminoacridine hydrochloride	Zaragozic acid A	Zaragozic acid A	Zaragozic acid A	Zaragozic acid A
B	Blasticidin S	Blasticidin S	Blasticidin S	Blasticidin S	Thioridazine hydrochloride	Thioridazine hydrochloride	Thioridazine hydrochloride	Thioridazine hydrochloride	Sodium Benzoate	Sodium Benzoate	Sodium Benzoate	Sodium Benzoate
C	Chlortetracycline hydrochloride	Chlortetracycline hydrochloride	Chlortetracycline hydrochloride	Chlortetracycline hydrochloride	Sodium Meta silicate	Sodium Meta silicate	Sodium Meta silicate	Sodium Meta silicate	Pentamidine Isethionate	Pentamidine Isethionate	Pentamidine Isethionate	Pentamidine Isethionate
D	6-Azauracil	6-Azauracil	6-Azauracil	6-Azauracil	Potassium Chromate	Potassium Chromate	Potassium Chromate	Potassium Chromate	Thialysine	Thialysine	Thialysine	Thialysine
E	a-Monothioglycerol	a-Monothioglycerol	a-Monothioglycerol	a-Monothioglycerol	EGTA	EGTA	EGTA	EGTA	Sodium Pyrophosphate decahydrate	Sodium Pyrophosphate decahydrate	Sodium Pyrophosphate decahydrate	Sodium Pyrophosphate decahydrate
F	Propiconazole	Propiconazole	Propiconazole	Propiconazole	Methyl Viologen dichloride	Methyl Viologen dichloride	Methyl Viologen dichloride	Methyl Viologen dichloride	Sodium Fluoride	Sodium Fluoride	Sodium Fluoride	Sodium Fluoride
G	Cisplatin	Cisplatin	Cisplatin	Cisplatin	Aluminum Sulfate	Aluminum Sulfate	Aluminum Sulfate	Aluminum Sulfate	Berberine chloride	Berberine chloride	Berberine chloride	Berberine chloride
H	Isoniazid	Isoniazid	Isoniazid	Isoniazid	Amphotericin B	Amphotericin B	Amphotericin B	Amphotericin B	Miconazole Nitrate	Miconazole Nitrate	Miconazole Nitrate	Miconazole Nitrate

Annexe E: Phenotype MicroArray panels

H. PM25 – Chemical Sensitivity

	1	2	3	4	5	6	7	8	9	10	11	12
A	Hydroxyurea	Hydroxyurea	Hydroxyurea	Hydroxyurea	Tobramycin	Tobramycin	Tobramycin	Tobramycin	Niaproof	Niaproof	Niaproof	Niaproof
B	b-Chloro-L-Alanine hydrochloride	b-Chloro-L-Alanine hydrochloride	b-Chloro-L-Alanine hydrochloride	b-Chloro-L-Alanine hydrochloride	Tetrazolium Violet	Tetrazolium Violet	Tetrazolium Violet	Tetrazolium Violet	Kanamycin Sulfate	Kanamycin Sulfate	Kanamycin Sulfate	Kanamycin Sulfate
C	4-Aminopyridine	4-Aminopyridine	4-Aminopyridine	4-Aminopyridine	Amitriptyline hydrochloride	Amitriptyline hydrochloride	Amitriptyline hydrochloride	Amitriptyline hydrochloride	Citric acid trisodium salt	Citric acid trisodium salt	Citric acid trisodium salt	Citric acid trisodium salt
D	Mechlorethamine hydrochloride	Mechlorethamine hydrochloride	Mechlorethamine hydrochloride	Mechlorethamine hydrochloride	Hygromycin B	Hygromycin B	Hygromycin B	Hygromycin B	5-Fluoro-2`-deoxy uridine	5-Fluoro-2`-deoxy uridine	5-Fluoro-2`-deoxy uridine	5-Fluoro-2`-deoxy uridine
E	Sodium Salicylate	Sodium Salicylate	Sodium Salicylate	Sodium Salicylate	Succinic acid	Succinic acid	Succinic acid	Succinic acid	Clomiphene Citrate	Clomiphene Citrate	Clomiphene Citrate	Clomiphene Citrate
F	Malic acid	Malic acid	Malic acid	Malic acid	Tartaric acid	Tartaric acid	Tartaric acid	Tartaric acid	Fumaric acid	Fumaric acid	Fumaric acid	Fumaric acid
G	5-Fluorocytosine	5-Fluorocytosine	5-Fluorocytosine	5-Fluorocytosine	Palladium(II) chloride	Palladium(II) chloride	Palladium(II) chloride	Palladium(II) chloride	Ibuprofen	Ibuprofen	Ibuprofen	Ibuprofen
H	Chloroquine	Chloroquine	Chloroquine	Chloroquine	Cinnamic acid	Cinnamic acid	Cinnamic acid	Cinnamic acid	Fluorouracil	Fluorouracil	Fluorouracil	Fluorouracil

Annexe F1: List of differentially expressed genes in T0KO1 strain.

orf19.1030, orf19.1075, orf19.1272, orf19.1287, orf19.1305, orf19.1335, orf19.1365, orf19.1430, orf19.1578, orf19.1584, orf19.1639, orf19.1835, orf19.1939, orf19.1958, orf19.2017, orf19.2038, orf19.2059, orf19.2168, orf19.246, orf19.2468, orf19.2478, orf19.2515, orf19.2638, orf19.270, orf19.2701, orf19.2770, orf19.2925, orf19.3107, orf19.3470, orf19.3475, orf19.3516, orf19.3547, orf19.3633, orf19.3665, orf19.3684, orf19.3689, orf19.3810, orf19.3910, orf19.3957, orf19.3970, orf19.4121, orf19.4125, orf19.4161, orf19.4283, orf19.4286, orf19.4479, orf19.4518, orf19.4530.1, orf19.4581, orf19.4642, orf19.478, orf19.4791, orf19.4814, orf19.4875, orf19.4964, orf19.4970, orf19.499, orf19.501, orf19.5038, orf19.5125, orf19.5169, orf19.518, orf19.5277orf19.5339 orf19.536 orf19.5413 orf19.5488 orf19.5517 orf19.552 orf19.5589 orf19.5611 orf19.5655 orf19.5784 orf19.581 orf19.5813 orf19.5862 orf19.5905 orf19.6090 orf19.6355 orf19.6360 orf19.6418 orf19.6589 orf19.6592 orf19.6637 orf19.6641 orf19.6723 orf19.6788 orf19.6866 orf19.687 orf19.6873.1 orf19.6950 orf19.6982 orf19.7011 orf19.7056 orf19.715 orf19.7160 orf19.7385 orf19.776, AAT22, ACO2, ALS1, AQY1, ARG4, ARG5,6, ARG8, ARP9, ASG7, ASR1, ASR3, AVT1, BMT3, BUD31, CAR2, CCC1, CDC6, CFL2, CHO2, CPA1, CSH1, CSI2, CTN3, DBP3, DBP7, DRS1, ECM42, ENG1, EXG2, FIL1, FMP45, FTH1, GAL10, GAP2, GAP6, GCA1, GCA2, GCV2, GUA1, GYP1, HAK1, HIS5, ICL1, IFD6, IHD2, KRE30, LAP3, LIP3, MEP2, MIS12, MUP1, NAG3, NIP7, NOG1, NOP10, OPT1, OYE22, OYE23, PCL2, PGA7, POL93, PSA2, PST1PST2 RBT5 RDI1 RPA190 RPC40 SHA3 SIK1 SKO1 SMM1 SNO1 SOD6 SOU1 SSP96 SUP35 TEF4 UGA1 WSC2 YVH1 ZCF4

Annexe F2: List of differentially expressed genes in T1 strain.

orf19.1007, orf19.102, orf19.1026, orf19.1034, orf19.1045, orf19.107, orf19.1075, orf19.1083, orf19.1087, orf19.1091, orf19.11, orf19.1107, orf19.1110, orf19.1117, orf19.1124.2, orf19.1163, orf19.1181, orf19.1190, orf19.1200, orf19.1201, orf19.1217, orf19.1240, orf19.1257, orf19.1267, orf19.1268, orf19.1277, orf19.1285, orf19.1287, orf19.1296, orf19.1303, orf19.1305, orf19.1307, orf19.1308, orf19.1310, orf19.1314, orf19.1326, orf19.1335, orf19.1344, orf19.1356, orf19.1365, orf19.1368, orf19.137, orf19.1381, orf19.1386, orf19.1388, orf19.1389, orf19.1393, orf19.1403, orf19.1404, orf19.1421, orf19.1426, orf19.1430, orf19.1433, orf19.1449, orf19.1484, orf19.1486, orf19.1491, orf19.1504, orf19.1507, orf19.1528, orf19.1532, orf19.1533, orf19.1539orf19.154 orf19.1544 orf19.1545 orf19.1547 orf19.1548 orf19.1549 orf19.1562 orf19.158 orf19.1582 orf19.1584 orf19.1588 orf19.1589 orf19.1590 orf19.1600 orf19.1610 orf19.1611 orf19.163 orf19.1632 orf19.1639 orf19.1646 orf19.1653 orf19.1654 orf19.1662 orf19.1664 orf19.1668 orf19.168 orf19.1687 orf19.1691 orf19.1697 orf19.1708 orf19.1723 orf19.1736 orf19.1737 orf19.175 orf19.1754 orf19.1765 orf19.1774 orf19.1780 orf19.1785 orf19.1789 orf19.1791 orf19.1796 orf19.1797 orf19.1802 orf19.1815 orf19.182 orf19.1825 orf19.1835 orf19.1841 orf19.185 orf19.1856 orf19.1857 orf19.1862 orf19.188 orf19.1889 orf19.190 orf19.1917 orf19.1952 orf19.1956 orf19.1959 orf19.1961 orf19.1964 orf19.1967 orf19.1985 orf19.200 orf19.2006 orf19.2009 orf19.2017 orf19.2018 orf19.2018.1 orf19.2038 orf19.2043 orf19.2045 orf19.2047 orf19.2048 orf19.2050 orf19.2067 orf19.2072 orf19.2089 orf19.2090 orf19.210 orf19.2106 orf19.2114 orf19.2115 orf19.2128 orf19.2132 orf19.2143 orf19.2163 orf19.2168 orf19.2175 orf19.2178 orf19.2191 orf19.22 orf19.2204 orf19.2214 orf19.223 orf19.2247 orf19.2249 orf19.2256 orf19.2259 orf19.2261 orf19.2262 orf19.2269 orf19.2275 orf19.2285 orf19.229 orf19.2309 orf19.2310 orf19.2312 orf19.2313 orf19.2314 orf19.2318.1 orf19.2319 orf19.2320 orf19.2335 orf19.2362 orf19.2363 orf19.2382 orf19.2384 orf19.2386 orf19.2389 orf19.2414 orf19.2426 orf19.2445 orf19.2457 orf19.2458 orf19.246 orf19.247 orf19.2478 orf19.2484 orf19.25 orf19.2500 orf19.2515 orf19.252 orf19.2521 orf19.2527 orf19.2529 orf19.2533.1 orf19.254 orf19.2542 orf19.2544 orf19.2547 orf19.2552 orf19.2563 orf19.2575 orf19.2591 orf19.2594 orf19.2604 orf19.2620 orf19.2629 orf19.2631 orf19.2638 orf19.2639 orf19.2639.1 orf19.264 orf19.2642 orf19.2657 orf19.2670 orf19.2671 orf19.268 orf19.2684 orf19.2686 orf19.2690 orf19.2691 orf19.2697 orf19.270 orf19.2701 orf19.2711 orf19.2720 orf19.2724 orf19.2726 orf19.2728 orf19.2733 orf19.2736 orf19.2739 orf19.276 orf19.2779 orf19.2781 orf19.2796 orf19.28 orf19.2812 orf19.2825 orf19.2826 orf19.2835 orf19.2836 orf19.2838 orf19.2847 orf19.2853 orf19.2866 orf19.2887 orf19.2890 orf19.2905 orf19.2917 orf19.2925 orf19.2930 orf19.2965.1 orf19.2966 orf19.2982 orf19.3004 orf19.3021 orf19.304 orf19.3048 orf19.3076 orf19.3088 orf19.3089

orf19.310 orf19.3100 orf19.3105 orf19.3114 orf19.3124 orf19.3130 orf19.3141 orf19.3156
orf19.3161 orf19.3167 orf19.3170 orf19.3177 orf19.318 orf19.3185 orf19.319 orf19.3204 orf19.321
orf19.3216 orf19.3218 orf19.3219 orf19.3248 orf19.3260 orf19.3261 orf19.3273 orf19.3281
orf19.3283 orf19.3285 orf19.3288.1 orf19.3295 orf19.3296 orf19.3302 orf19.3303 orf19.3304
orf19.3310 orf19.3317 orf19.332 orf19.3325 orf19.3333 orf19.3338 orf19.3342 orf19.3352
orf19.3357 orf19.3362 orf19.3364 orf19.337 orf19.3378 orf19.3393 orf19.34 orf19.3406 orf19.341
orf19.3418 orf19.3425 orf19.3427 orf19.3442 orf19.3444 orf19.3447 orf19.3448 orf19.345
orf19.3461 orf19.3463 orf19.3470 orf19.3479 orf19.3480 orf19.3481 orf19.3483 orf19.3501
orf19.3508 orf19.3512 orf19.3515 orf19.3516 orf19.3537 orf19.3539 orf19.354 orf19.3544
orf19.3547 orf19.3556 orf19.3559 orf19.3577.1 orf19.3582 orf19.3586 orf19.3604 orf19.3607
orf19.3613 orf19.3615 orf19.3624 orf19.3625 orf19.3633 orf19.3643 orf19.3649 orf19.3665
orf19.3667 orf19.3671 orf19.3679 orf19.3684 orf19.3686 orf19.3704 orf19.371 orf19.3720
orf19.3755 orf19.3759 orf19.376 orf19.3767 orf19.3778 orf19.3780 orf19.3781 orf19.3788
orf19.3793 orf19.3798 orf19.3804 orf19.3817 orf19.3820 orf19.3854 orf19.3869 orf19.3881
orf19.3885 orf19.3897 orf19.3902 orf19.3908 orf19.3910 orf19.3920 orf19.3921 orf19.3922
orf19.3928 orf19.3932 orf19.3938 orf19.3970 orf19.3976 orf19.398 orf19.3984 orf19.3991
orf19.4005 orf19.4007 orf19.4013 orf19.4017 orf19.4018 orf19.4031 orf19.4090 orf19.4091
orf19.4092 orf19.4096 orf19.4097 orf19.4105 orf19.4106 orf19.4107 orf19.411 orf19.4115
orf19.4121 orf19.4122 orf19.4125 orf19.4132 orf19.414 orf19.4142 orf19.4159 orf19.4160
orf19.4161 orf19.417 orf19.4170 orf19.4173 orf19.4182 orf19.4185 orf19.4190 orf19.4194
orf19.4195 orf19.4227 orf19.423 orf19.4240 orf19.4241 orf19.4243 orf19.4269 orf19.427
orf19.4273 orf19.428 orf19.4283 orf19.4287 orf19.429 orf19.4293 orf19.4306 orf19.4326
orf19.4342 orf19.4355 orf19.4358 orf19.4362 orf19.4365 orf19.4366 orf19.4368 orf19.4375
orf19.4391 orf19.4399 orf19.4409 orf19.443 orf19.4435 orf19.4436 orf19.4445 orf19.4449
orf19.4450.1 orf19.4459 orf19.4465 orf19.4479 orf19.448 orf19.4488 orf19.449 orf19.4498
orf19.4503 orf19.4504 orf19.4507 orf19.4515 orf19.4518 orf19.4523 orf19.4530.1 orf19.4538
orf19.4539 orf19.4550 orf19.4557 orf19.4563 orf19.4569 orf19.457 orf19.4570 orf19.4571
orf19.4580 orf19.4581 orf19.4589 orf19.4594 orf19.4595 orf19.4600 orf19.4611 orf19.4612
orf19.4614 orf19.4617 orf19.4626 orf19.4633 orf19.4634 orf19.4642 orf19.4652 orf19.4657
orf19.4665 orf19.4666 orf19.4668 orf19.467 orf19.4675 orf19.4686 orf19.4690 orf19.4705
orf19.4706 orf19.4713 orf19.4721 orf19.4723 orf19.4730 orf19.4734 orf19.4735 orf19.4738
orf19.474 orf19.4749 orf19.475 orf19.4750 orf19.4756 orf19.4760 orf19.4764 orf19.478 orf19.4792
orf19.4793 orf19.4795 orf19.4801 orf19.4805 orf19.4828 orf19.4843 orf19.4850 orf19.4886
orf19.4889 orf19.4895 orf19.4898 orf19.4911 orf19.4912 orf19.4913 orf19.4914 orf19.4916
orf19.4918 orf19.4931 orf19.494 orf19.4940 orf19.4942 orf19.4948 orf19.496 orf19.4963
orf19.4970 orf19.499 orf19.50 orf19.500 orf19.501 orf19.5012 orf19.5020 orf19.5022 orf19.5026
orf19.5030 orf19.5037 orf19.5038 orf19.5039 orf19.5043 orf19.5049 orf19.5078 orf19.5085
orf19.5090 orf19.510 orf19.5103 orf19.5125 orf19.5126 orf19.5136 orf19.5141 orf19.5143
orf19.515 orf19.5156 orf19.5158 orf19.5161 orf19.5169 orf19.5175 orf19.5177 orf19.5201
orf19.5203 orf19.5205 orf19.5206 orf19.5207 orf19.5210 orf19.5216 orf19.5220 orf19.5221
orf19.5238 orf19.5239 orf19.524 orf19.5253 orf19.5266 orf19.5270 orf19.5277 orf19.5278
orf19.5279 orf19.5282 orf19.5287 orf19.5290 orf19.5293 orf19.5295 orf19.5326 orf19.5346
orf19.5356 orf19.5363 orf19.5364 orf19.5365 orf19.5370 orf19.5391 orf19.540 orf19.5401
orf19.5409 orf19.541 orf19.5431 orf19.5450 orf19.5451 orf19.5455 orf19.5457 orf19.5459
orf19.547 orf19.5483 orf19.549 orf19.5490 orf19.5504 orf19.551 orf19.5514 orf19.5517 orf19.5518
orf19.552 orf19.5524 orf19.5525 orf19.5527 orf19.5532 orf19.5541 orf19.5565 orf19.5566
orf19.557 orf19.5576 orf19.5587 orf19.5589 orf19.5608 orf19.5611 orf19.5619 orf19.5620
orf19.5625 orf19.5627 orf19.5642 orf19.5648 orf19.5665 orf19.5678 orf19.5681 orf19.5686
orf19.5698 orf19.5704 orf19.5722 orf19.5763 orf19.577 orf19.5772 orf19.5782 orf19.5783
orf19.5784 orf19.5799 orf19.580 orf19.5802 orf19.5808 orf19.5812 orf19.5813 orf19.5824
orf19.5834 orf19.5838 orf19.5840 orf19.5841 orf19.5843 orf19.5847 orf19.585 orf19.5862
orf19.5863 orf19.588 orf19.5892 orf19.5905 orf19.5915 orf19.5920 orf19.5932 orf19.5943
orf19.5961 orf19.5965 orf19.5980 orf19.5987 orf19.5991 orf19.6007 orf19.6016 orf19.6017
orf19.6022 orf19.6024 orf19.6030 orf19.6045 orf19.6048 orf19.605 orf19.6055 orf19.6065
orf19.6066 orf19.6071 orf19.6082 orf19.6110 orf19.6113 orf19.6117 orf19.6136 orf19.6137
orf19.6147 orf19.6152 orf19.6175 orf19.6194 orf19.6200 orf19.6225.1 orf19.6234 orf19.6247
orf19.6291 orf19.6297 orf19.6311 orf19.6316.4 orf19.6328 orf19.633 orf19.6348 orf19.6355
orf19.6357 orf19.6360 orf19.6371 orf19.639 orf19.6392 orf19.6396 orf19.6418 orf19.6431

orf19.6434 orf19.6443 orf19.6448 orf19.647.3 orf19.6477 orf19.6484 orf19.6487 orf19.649 orf19.6499 orf19.6503 orf19.6528 orf19.6532 orf19.6553 orf19.6558 orf19.6559 orf19.6586 orf19.6589 orf19.6637 orf19.6641 orf19.6653 orf19.6676 orf19.6679 orf19.6688 orf19.6705 orf19.6720 orf19.6723 orf19.673 orf19.6730 orf19.6736 orf19.6739 orf19.6741 orf19.6751 orf19.6753 orf19.6756 orf19.6758 orf19.6766 orf19.6769 orf19.6786 orf19.6793 orf19.680 orf19.6805 orf19.6809 orf19.6818 orf19.6828 orf19.6828.1 orf19.6829 orf19.6830 orf19.6832 orf19.6838 orf19.6840 orf19.6845 orf19.6852 orf19.6853 orf19.6858 orf19.6862 orf19.6873.1 orf19.6879 orf19.688 orf19.6886 orf19.6893 orf19.6899 orf19.6903 orf19.6905 orf19.6907 orf19.6917 orf19.692 orf19.6921 orf19.6933 orf19.6952 orf19.6968 orf19.6973 orf19.6982 orf19.6983 orf19.6984 orf19.6989 orf19.6990 orf19.6992 orf19.6995 orf19.6996 orf19.7010 orf19.7011 orf19.7012 orf19.7013 orf19.702 orf19.7023 orf19.7028 orf19.703 orf19.7041 orf19.7042 orf19.7044 orf19.7056 orf19.7063 orf19.7074 orf19.7085 orf19.7088 orf19.7091 orf19.7096 orf19.7104 orf19.7106 orf19.7127.1 orf19.7130 orf19.715 orf19.7153 orf19.7159 orf19.7160 orf19.719 orf19.7192 orf19.7193 orf19.7196 orf19.7197 orf19.7199 orf19.7200 orf19.7204 orf19.7209 orf19.7210 orf19.7212 orf19.7215 orf19.7223 orf19.7225 orf19.7227 orf19.725 orf19.7250 orf19.7252 orf19.7254 orf19.7271 orf19.7288 orf19.7291 orf19.73 orf19.7304 orf19.7307 orf19.7310 orf19.7311 orf19.7328 orf19.733 orf19.7344 orf19.7347 orf19.7361 orf19.7370 orf19.7385 orf19.7386 orf19.7392 orf19.7396 orf19.7398 orf19.7403 orf19.7404 orf19.7405 orf19.7406 orf19.7409.1 orf19.7422 orf19.7425 orf19.7426 orf19.7427 orf19.7437 orf19.7445 orf19.7457 orf19.7459 orf19.7478 orf19.7488 orf19.7494 orf19.7504 orf19.7507 orf19.7511 orf19.7519 orf19.7522 orf19.7531 orf19.7539.1 orf19.7545 orf19.7546 orf19.7547 orf19.7554 orf19.7566 orf19.7567 orf19.7590 orf19.7593 orf19.7596 orf19.7598 orf19.7601 orf19.7617 orf19.7618 orf19.7620 orf19.7621 orf19.7624 orf19.7648 orf19.7663 orf19.7675 orf19.773 orf19.775 orf19.776 orf19.792 orf19.804 orf19.808 orf19.809 orf19.822 orf19.827 orf19.828 orf19.841 orf19.846 orf19.863 orf19.867 orf19.904 orf19.915 orf19.92 orf19.921 orf19.93 orf19.932 orf19.94 orf19.993 orf19.996, AAH1, AAP1, AAT1, ABC1, ABP1, ABP140, ADH3, ADH5, AHP1, ALD6, ALG11, ALG2, ALK2, ALK6, ALK8, ALO1, ALS5, ALS9, AMO2, AOX2, APA2, APL5, AQY1, ARA1, ARE2, ARF3, ARG1, ARG11, ARG3, ARG5,6, ARG8, ARO7, ARP8, ARX1, ASE1, ASR2, ASR3, ATC1, ATF1, ATG9, AUT7, AXL1, AXL2, AYR2, BCY1, BEM2, BMT3, BMT6, BNR1, BPH1, BRG1, BUB3, BUD16, BUD21, BUD22, BUD23, BUD31, BZZ1, CAF16, CAN1, CAN2, CAR1, CAR2CAT2 CBF1 CCC1 CCN1 CCP1 CCT7 CDC34 CDC46 CDC54 CDR4 CFL2 CFL4 CFL5 CHO2 CHR1 CHS1 CHS2 CHS3 CHS5 CHT1 CIC1 CIP1 CIRT4B CKA1 CKA2 CLG1 CMK1 CNS1 COX15 CPA1 CPA2 CPY1 CRH11 CRH12 CRP1 CRZ1 CRZ2 CSI2 CSP37 CST5 CTA1 CTA2 CTA24 CTA26 CTA7 CTA8 CTA9 CTF1 CTF18 CTF8 CTM1 CTN1 CTN3 CTP1 CUP9 CYB2 CYS3 CYT2 DAK2 DAL1 DAP1 DBP2 DBP3 DBP7 DBP8 DEF1 DES1 DFG16 DFI1 DIM1 DIP2 DIP5 DLD1 DOA1 DOT4 DPM1 DPP1 DRG1 DRS1 DUR1,2 EAF3 EAF6 ECI1 ECM18 ECM21 ECM3 ECM38 ECM39 ECM42 ELF1 ELP3 ENP1 ENP2 ERB1 ERG1 ESA1 ESS1 EST1 EXG2 EXO1 EXO70 FAD3 FAL1 FAT1 FAV1 FCA1 FCY2 FCY23 FCY24 FDH1 FET3 FET34 FGR17 FGR23 FGR32 FGR39 FGR50 FIL2 FLC3 FLO8 FLO9 FLU1 FMO1 FMP45 FOX2 FOX3 FPG1 FRE10 FRE9 FRP1 FRP2 FTH2 FUM12 GAL7 GAP2 GAT1 GCA1 GCA2 GCD2 GCD7 GCN20 GCY1 GDB1 GDH2 GDS1 GIG1 GIN4 GIT4 GLC3 GLE1 GLK4 GLN3 GLT1 GNP3 GPD1 GPD2 GPI13 GPI7 GPM2 GPR1 GRE3 GRP2 GRR1 GST2 GSY1 GUT1 GYP1 GYP7 GZF3 HAK1 HAP3 HAP41 HBR3 HCA4 HCM1 HEM1 HEM13 HEM14 HGT10 HGT13 HGT14 HGT16 HGT17 HGT18 HGT5 HGT6 HIS3 HIS5 HMT1 HNM1 HNM3 HNT1 HNT2 HOC1 HOL1 HRK1 HRT2 HSP31 HST1 HST2 HXT5 HYM1 HYR1 ICL1 IDI1 IDP2 IFA14 IFA21 IFC1 IFC3 IFD6 IFE2 IFF11 IFF4 IFF6 IFF9 IFG3 IFM3 IFR2 IFU5 IHD1 IML2 IMP4 INP51 IPT1 ISA1 ISN1 ISU1 JEN2 JIP5 KAR3 KIN2 KIS2 KRE62 KRR1 KTI12 LAP3 LAS1 LCB2 LCB4 LDG3 LEM3 LIG1 LIG4 LIP1 LIP5 LIP9 LTV1 LYS2 MAF1 MAK16 MAK5 MAL2 MAL31 MAS2 MBP1 MCD1 MCM2 MCM3 MCM6 MCT1 MDH1-3 MDM34 MEF2 MEP1 MET10 MET13 MET14 MET16 MET3 MEU1 MEX67 MGM101 MKK2 MNN1 MNN10 MNN22 MPP10 MRP17 MRPL10 MRPL3 MRPL40 MRPL6 MRPL8 MRPS9 MRR1 MRT4 MSF1 MSH6 MSS116 MTO1 MTR10 MTW1 MUM2 MUQ1 MVB12 NAG3 NAG6 NAN1 NCE103 NCR1 NDH51 NEP1 NIP7 NMA111 NMD3 NMT1 NOC4 NOG2 NOP14 NRG1 NSA1 NSA2 NTG1 NTH1 NUF2 NUP OAC1 OGG1 OPI1 OPT1 OPT4 OPT5 OPT7 OPT9 ORM1 OSM1 OSM2 OXR1 OYE23 OYE32 PAM16 PBR1 PCK1 PCL2 PCL7 PCT1 PDK2 PDR6 PDS5 PDX1 PDX3 PEX11 PEX13 PEX14 PEX4 PEX5 PFK1 PFK26 PGA10 PGA13 PGA23 PGA27 PGA28 PGA31 PGA32 PGA37 PGA57 PGA6 PGA7 PGM2 PHA2 PHHB PHO113 PHO15 PHO23 PLB1 PNG2 POL2 POL32 POP3 POP4 POS5 POT1 POT1-2 POX18 PPR1 PPT1 PRB1 PRC2 PRD1 PRI2 PRM1 PRO1 PRO2 PRP5 PRS1 PRS5 PRT1 PSA2 PSF2 PST1 PST2 PSY2

PTC4 PTC8 PTH2 PUF3 PUT1 PUT2 PWP2 PXP2 PYC2 QDR1 RAD16 RAD52 RAD54 RAT1 RBD1 RBF1 RBT4 RCL1 REI1 RFC4 RFX2 RGD3 RHR2 RIM11 RIM13 RIM2 RIM20 RIM8 RIO2 RIX7 RKI1 RLI1 RLM1 RLP24 RMT2 RNH35 RNR22 RNR3 ROD1 RPA12 RPA190 RPC19 RPC40 RPC53 RPD31 RPF1 RPG1A RPL35 RPL7 RPN6 RPO41 RRN11 RRN3 RRP15 RRP6 RRP9 RRS1 RSN1 RSR1 RTA2 RTA3 RTA4 RTF1 RVS162 SAP3 SAP99 SAS3 SCH9 SCS7 SCT1 SDA1 SDC1 SDH4 SDS22 SEC2 SEC20 SEF1 SEN1 SEN2 SEO1 SET3 SFC1 SFL1 SFU1 SGD1 SIR2 SKN7 SLA2 SLD1 SLD5 SLP2 SMC2 SMC3 SMM1 SMP3 SMT3 SNM1 SNQ2 SOF1 SOL3 SPB4 SPE2 SPO72 SPR28 SPT10 SRB8 SRP40 SSP96 STF2 STP3 STP4 SUR2 SVF1 SYG1 SYS3 TAF145 TAF4 TBF1 TCC1 TEA1 TES1 TFB3 TFS1 THR1 TIF5 TLG2 TLO1 TLO10 TLO11 TLO13 TLO34 TLO4 TLO5 TLO7 TLO9 TNA1 TOR1 TPK1 TPM2 TPO2 TPO4 TPS1 TPS2 TPS3 TRM1 TRM2 TRP3 TRP4 TSC11 TSR1 TSR2 TYR1 UBA4 UBC8 UGA1 UGA11 UGA4 UGA6 ULP1 ULP2 UPC2 URA1 URA2 URA3 URA5 URK1 USO6 UTP13 UTP18 UTP20 UTP21 UTP22 UTP4 UTP8 UTP9 VPS17 VPS2 VPS20 VPS21 VPS35 VTC4 WRS1 WSC4 XOG1 XYL2 YAH1 YDC1 YEA4 YHB5 YHM1 YKU80 YML6 YOR1 YOX1 YPS7 YTA6 YTM1 YVH1 ZCF17 ZCF2 ZCF23 ZCF24 ZCF25 ZCF35 ZCF37

Annexe F3: List of differentially expressed genes in T1KO1 strain.

orf19.1007, orf19.102, orf19.1030, orf19.1034, orf19.1043, orf19.1058, orf19.107, orf19.1087, orf19.11, orf19.1106, orf19.1107, orf19.1117, orf19.1200, orf19.1253, orf19.1257, orf19.1267, orf19.1272, orf19.1281, orf19.1318, orf19.1335, orf19.1344, orf19.1356, orf19.136, orf19.1365, orf19.1368, orf19.1381, orf19.1383, orf19.1388, orf19.1389, orf19.1395, orf19.1404, orf19.1430, orf19.1433, orf19.1444, orf19.1449, orf19.1473, orf19.1486, orf19.1532, orf19.1533, orf19.1539, orf19.154, orf19.1544, orf19.1547, orf19.1562, orf19.1567, orf19.1577, orf19.1584, orf19.1600, orf19.1608, orf19.1639, orf19.1641, orf19.1646, orf19.1654, orf19.1668, orf19.1679, orf19.1691, orf19.1697, orf19.1723, orf19.1729, orf19.1736, orf19.1737, orf19.1754, orf19.1765orf19.1771 orf19.1774 orf19.1778 orf19.1796 orf19.1799 orf19.1802 orf19.1815 orf19.1856 orf19.1857 orf19.1867 orf19.187 orf19.188 orf19.1888 orf19.1897 orf19.190 orf19.1939 orf19.1952 orf19.1961 orf19.200 orf19.2018.1 orf19.2024 orf19.2038 orf19.2059 orf19.2071 orf19.2090 orf19.2114 orf19.2125 orf19.2175 orf19.2180 orf19.22 orf19.2204 orf19.223 orf19.2247 orf19.2249 orf19.2260 orf19.2263 orf19.2269 orf19.2285 orf19.2310 orf19.2312 orf19.2335 orf19.2366 orf19.2386 orf19.2392 orf19.2414 orf19.2436 orf19.2445 orf19.2457 orf19.2458 orf19.246 orf19.2468 orf19.247 orf19.2484 orf19.2515 orf19.2517 orf19.2529 orf19.254 orf19.2542 orf19.2552 orf19.2575 orf19.258 orf19.2607 orf19.2620 orf19.263.1 orf19.2638 orf19.264 orf19.2684 orf19.2691 orf19.2697 orf19.2701 orf19.2720 orf19.2724 orf19.2726 orf19.2736 orf19.2769 orf19.2770 orf19.278 orf19.2796 orf19.2812 orf19.2825 orf19.2826 orf19.2835 orf19.2844 orf19.2850 orf19.2853 orf19.2863.1 orf19.2866 orf19.2870 orf19.2930 orf19.2978 orf19.3009 orf19.304 orf19.3088 orf19.310 orf19.3105 orf19.3124 orf19.3141 orf19.3156 orf19.3216 orf19.3222 orf19.3248 orf19.3281 orf19.3283 orf19.3288.1 orf19.3296 orf19.3302 orf19.3304 orf19.332 orf19.3325 orf19.3338 orf19.3368 orf19.338 orf19.3393 orf19.3399 orf19.34 orf19.3404 orf19.3406 orf19.341 orf19.3418 orf19.3432 orf19.344 orf19.3444 orf19.3447 orf19.3448 orf19.345 orf19.3456 orf19.3469 orf19.3470 orf19.3473 orf19.3483 orf19.3512 orf19.3536 orf19.3544 orf19.3546 orf19.3577.1 orf19.3604 orf19.3607 orf19.3610 orf19.3625 orf19.3643 orf19.3706 orf19.371 orf19.3759 orf19.3767 orf19.3778 orf19.3780 orf19.3781 orf19.3798 orf19.3804 orf19.3810 orf19.3820 orf19.3828 orf19.3831 orf19.384 orf19.3848 orf19.3869 orf19.3871 orf19.3897 orf19.3908 orf19.3910 orf19.3914 orf19.392 orf19.3920 orf19.3924 orf19.3926 orf19.3991 orf19.4007 orf19.4031 orf19.4062 orf19.4070 orf19.4096 orf19.4105 orf19.4106 orf19.4115 orf19.4121 orf19.4125 orf19.4128 orf19.413 orf19.4142 orf19.415 orf19.4161 orf19.417 orf19.4170 orf19.4172 orf19.4173 orf19.4190 orf19.4203 orf19.4227 orf19.423 orf19.4240 orf19.4243 orf19.4246 orf19.4250 orf19.427 orf19.4273 orf19.4283 orf19.4287 orf19.4294 orf19.4306 orf19.4316 orf19.4325 orf19.4326 orf19.4342 orf19.4355 orf19.4362 orf19.4365 orf19.4368 orf19.4391 orf19.4399 orf19.4407 orf19.4439 orf19.4439.1 orf19.4445 orf19.4476 orf19.4479 orf19.448 orf19.4498 orf19.4503 orf19.4504 orf19.4507 orf19.4515 orf19.4518 orf19.4523 orf19.4539 orf19.4550 orf19.4557 orf19.4563 orf19.4569 orf19.4571 orf19.4577 orf19.4580 orf19.4581 orf19.4594 orf19.4612 orf19.4614 orf19.4621 orf19.4634 orf19.4642 orf19.4652 orf19.4666 orf19.4668 orf19.4675 orf19.4686 orf19.4706 orf19.4735 orf19.4738 orf19.4749 orf19.4756 orf19.4760 orf19.4805 orf19.4812 orf19.4818 orf19.4828 orf19.4835

orf19.4843 orf19.4873 orf19.4875 orf19.4880 orf19.4898 orf19.4912 orf19.4913 orf19.4929 orf19.494 orf19.496 orf19.499 orf19.5017 orf19.5020 orf19.5022 orf19.5026 orf19.5030 orf19.5037 orf19.5038 orf19.5043 orf19.5090 orf19.5103 orf19.5125 orf19.5136 orf19.5140 orf19.5141 orf19.5156 orf19.5169 orf19.5195 orf19.5203 orf19.5206 orf19.5207 orf19.5210 orf19.5216 orf19.5221 orf19.5225 orf19.5250 orf19.5256 orf19.5277 orf19.5278 orf19.5279 orf19.5290 orf19.5293 orf19.5326 orf19.5340 orf19.5346 orf19.5350 orf19.5356 orf19.5364 orf19.5365 orf19.5406 orf19.5418 orf19.5431 orf19.5451 orf19.5455 orf19.547 orf19.5490 orf19.5514 orf19.5525 orf19.5536 orf19.555 orf19.5564 orf19.5573 orf19.5587 orf19.5589 orf19.5608 orf19.5611 orf19.5619 orf19.5626 orf19.5627 orf19.5646 orf19.5648 orf19.5678 orf19.5704 orf19.5722 orf19.5763 orf19.5772 orf19.5783 orf19.5784 orf19.5802 orf19.5808 orf19.5812 orf19.5813 orf19.5824 orf19.5843 orf19.5871 orf19.5879 orf19.588 orf19.5915 orf19.5932 orf19.5953 orf19.5961 orf19.5978 orf19.5980 orf19.5987 orf19.6013 orf19.6017 orf19.6045 orf19.605 orf19.6055 orf19.607 orf19.6079 orf19.6080 orf19.6090 orf19.6143 orf19.6147 orf19.6175 orf19.6180 orf19.6184 orf19.6195 orf19.6227 orf19.6234 orf19.6247 orf19.6260 orf19.6291 orf19.6306 orf19.633 orf19.6350 orf19.6357 orf19.639 orf19.6403 orf19.6408 orf19.6432 orf19.6435 orf19.6448 orf19.6477 orf19.6484 orf19.6499 orf19.6530 orf19.6539 orf19.6558 orf19.6559 orf19.6586 orf19.6599.1 orf19.6637 orf19.664 orf19.6678 orf19.6679 orf19.6681 orf19.6688 orf19.6704 orf19.6705 orf19.6720 orf19.673 orf19.6740 orf19.6751 orf19.6753 orf19.6755 orf19.6758 orf19.6769 orf19.6786 orf19.6793 orf19.6809 orf19.6818 orf19.6832 orf19.6840 orf19.6879 orf19.6888 orf19.6893 orf19.6899 orf19.69 orf19.69.2 orf19.6903 orf19.6917 orf19.692 orf19.6921 orf19.6973 orf19.6982 orf19.6989 orf19.699 orf19.6992 orf19.7009 orf19.7010 orf19.7013 orf19.7020 orf19.7029 orf19.7044 orf19.7056 orf19.7057 orf19.7058 orf19.7073 orf19.7074 orf19.7078 orf19.7088 orf19.7091 orf19.7096 orf19.7098 orf19.7104 orf19.7106 orf19.7118 orf19.715 orf19.7153 orf19.7184 orf19.719 orf19.7192 orf19.7202 orf19.7244 orf19.725 orf19.7250 orf19.7271 orf19.7288 orf19.7290 orf19.7291 orf19.73 orf19.730 orf19.7301 orf19.7306 orf19.7307 orf19.7325 orf19.7330 orf19.7344 orf19.7347 orf19.7396 orf19.7405 orf19.7410 orf19.7426 orf19.7433 orf19.7437 orf19.7445 orf19.7459 orf19.7488 orf19.7494 orf19.7499 orf19.7507 orf19.7512 orf19.7531 orf19.7546 orf19.7589 orf19.7598 orf19.7601 orf19.7618 orf19.7621 orf19.7632 orf19.7642 orf19.7648 orf19.7662 orf19.775 orf19.804 orf19.808 orf19.809 orf19.812 orf19.822 orf19.841 orf19.846 orf19.851 orf19.880 orf19.929 orf19.932 orf19.952 orf19.967 orf19.980 orf19.996, AAH1, AAP1, AAT1, ABP140, ACB1, ACO2, ADE6, AFG1, AFT2, ALD6, ALK2, ALS9, AOX1, AOX2, APA2, AQY1, ARA1, ARG1, ARG11, ARG3, ARG4, ARG5,6, ARG8, ARH2, ARL3, ARO10, ARO3, ARX1, ASE1, ASG7, ASH2, ASR1, ATC1, ATG9, AVT1, AYR2, BEM3, BMS1, BMT3, BRG1, BUB2, BUD21, BUD23, BUL1, BUR2, CAC2, CAF16, CAN1, CAN2, CAR1, CAR2, CAT2, CCC1, CCT7, CDC34, CDG1, CEK1, CFL1, CFL2, CFL4, CFL5, CHA1, CHS1CHS5 CHT1 CHT3 CIC1 CIP1 CMK1 CNS1 CNT COX9 CPA1 CPA2 CRD2 CRH11 CRL1 CRP1 CRZ1 CRZ2 CSH1 CSI2 CTA1 CTA2 CTA8 CTF18 CTN1 CTN3 CTP1 CUP2 CUP9 CYB2 CYS3 DAL8 DAP1 DAP2 DBP7 DBP8 DDR48 DFG10 DIM1 DIP2 DIT2 DJP1 DLD1 DPP2 DRS1 DUR1,2 DUR4 EAF3 ECE1 ECI1 ECM18 ECM3 ECM331 ECM42 ELF1 ELP3 ENG1 ENP1 ENP2 ERG1 ERG3 EST1 EXM2 FAA21 FAD3 FAV1 FBP1 FCA1 FCY2 FCY21 FCY23 FET33 FET34 FGR17 FGR23 FGR50 FGR51 FIL1 FMO2 FMP45 FOX2 FOX3 FRE10 FRE9 FTH1 GAL10 GAL102 GAL4 GAL7 GAP1 GAP2 GAT1 GCD7 GCN5 GCV1 GCV2 GDH2 GDH3 GDS1 GFA1 GLC3 GLE1 GLG2 GNP1 GPD1 GPD2 GPI7 GPM2 GPR1 GPX1 GRE3 GRX1 GTT13 GYP1 GZF3 HAK1 HAP43 HAP5 HBR3 HEM1 HEM14 HGT12 HGT13 HGT14 HGT16 HGT6 HIS3 HIS4 HIS5 HNM1 HNM3 HNM4 HNT2 HOC1 HOL1 HOL4 HRK1 HRT2 HSP31 HST1 HWP1 HYR1 ICL1 IDI1 IDP2 IFA21 IFC1 IFC3 IFD6 IFF11 IFF6 IFG3 IFM3 IFR1 IFR2 IHD1 IHD2 IMP4 INO1 IPT1 JEN1 JEN2 JIP5 KGD1 KIN2 KIS2 KRR1 KTI12 LAP3 LCB2 LDG3 LEM3 LEU3 LIG1 LIG4 LIP5 LIP9 LTV1 MAC1 MBP1 MCD4 MCM2 MCT1 MDH1-3 MDR1 MEP1 MET10 MET14 MET15 MET16 MET3 MEX67 MIF2 MIG1 MIS12 MMS21 MNN1 MPP10 MRF1 MRPL33 MRR1 MTW1 MUP1 NAN1 NCE103 NEP1 NIP7 NOC4 NOG2 NOP14 NOP5 NPL3 NRG1 NSA1 NTH1 NUP OAC1 OGG1 OPT1 OPT4 OPT7 OPT9 OSM2 OXR1 OYE22 OYE23 PCL2 PCL7 PCT1 PDC12 PDS5 PDX1 PEX11 PEX13 PEX5 PFK26 PGA1 PGA10 PGA13 PGA26 PGA32 PGA43 PHA2 PHO114 PHO23 PHO87 PHO89 PHR1 PIF1 PIR1 PLB1 POL93 POP3 POP4 POS5 POT1 POX18 PRI2 PRM1 PRN2 PRN3 PRO1 PRO2 PRR2 PRS1 PRS5 PSP1 PTC5 PTH2 PTK2 PUT1 PUT2 PUT4 PWP2 QDR1 RAD16 RAD54 RAD57 RAS2 RBD1 RBF1 RBT4 RBT7 RCE1 RCL1 REI1 RFC4 RFX2 RIM11 RIM2 RLM1 RME1 RNR22 RNR3 RPA12 RPC19 RPD31 RRN3 RRP8 RRP9 RRS1 RSN1 RSR1 RTA2 RTA3 RTF1 RTT109 RXT3 SAC7 SAM2 SAP2 SAP6 SAS2 SAS3 SCW11 SDH4 SEC2 SEC20 SEC6 SEF1

SEN2 SET3 SFL1 SFU1 SGS1 SHA3 SHE9 SIK1 SIR2 SIT1 SKO1 SLD1 SMC2 SMC4 SMF12 SMM1 SMP3 SOD1 SOF1 SOU1 SPO11 SPO72 SPT10 SST2 STP3 STP4 SUA72 SUI3 SUL2 SUP35 SUR2 SYG1 SYS3 TAF145 TEA1 TERT TES1 TFB3 TLG2 TLO1 TLO34 TPO4 TPS1 TPS2 TPS3 TRM2 TRP3 TRP4 TRP99 TSC11 TSR1 TSR2 UBA4 UBC8 UGA1 UGA4 ULP1 ULP2 ULP3 UPC2 URA1 URA5 USO5 UTP13 UTP21 UTP4 UTP8 UTP9 VPH1 VTC3 WHI3 WOR1 WSC4 XYL2 YEA4 YHB5 YOX1 YPS7 YVH1 ZCF18 ZCF20 ZCF22 ZCF24 ZCF25 ZCF35 ZCF37 ZCF4 ZRT1 ZUO1

Annexe F4: List of differentially expressed genes in T2 strain.

orf19.1005, orf19.1007, orf19.1009, orf19.1026, orf19.1029, orf19.1030, orf19.104, orf19.1058, orf19.107, orf19.1075, orf19.1083, orf19.1087, orf19.11, orf19.1106, orf19.1107, orf19.1110, orf19.1114, orf19.1117, orf19.1122, orf19.1124.2, orf19.1146, orf19.1162.1, orf19.1167, orf19.1178, orf19.1190, orf19.1196, orf19.1200, orf19.1253, orf19.1267, orf19.1268, orf19.1272, orf19.1277, orf19.1281, orf19.1296, orf19.1305, orf19.1306, orf19.1318, orf19.132, orf19.1326, orf19.1335, orf19.1340, orf19.1343, orf19.1344, orf19.1365, orf19.1371, orf19.1383, orf19.1386, orf19.1388, orf19.1389, orf19.1395, orf19.1404, orf19.1424, orf19.1426, orf19.1430, orf19.1433, orf19.1449, orf19.1461, orf19.1473, orf19.1480, orf19.1484, orf19.1491, orf19.1504, orf19.1522orf19.1535 orf19.1539 orf19.154 orf19.1544 orf19.1547 orf19.1562 orf19.1578 orf19.1582 orf19.1584 orf19.1604 orf19.1608 orf19.1610 orf19.1626 orf19.1632 orf19.1639 orf19.1654 orf19.1678 orf19.1679 orf19.1682 orf19.1697 orf19.1708 orf19.1723 orf19.1728 orf19.1735 orf19.1736 orf19.1737 orf19.1745 orf19.1754 orf19.1764 orf19.1771 orf19.1772 orf19.1774 orf19.1778 orf19.1780 orf19.1789 orf19.1794 orf19.1796 orf19.1797 orf19.1802 orf19.1808 orf19.1852 orf19.1856 orf19.1857 orf19.1867 orf19.1871 orf19.1878 orf19.1888 orf19.1897 orf19.1938 orf19.1939 orf19.1940 orf19.1943 orf19.1948 orf19.1950 orf19.1959 orf19.1967 orf19.1968 orf19.1991 orf19.200 orf19.2007 orf19.2016 orf19.2019 orf19.2024 orf19.2030 orf19.2037 orf19.2038 orf19.2041 orf19.2042 orf19.2047 orf19.2048 orf19.2049 orf19.2050 orf19.2059 orf19.206 orf19.2066.1 orf19.2076 orf19.2088 orf19.2090 orf19.2111 orf19.2124 orf19.2125 orf19.2128 orf19.2143 orf19.216 orf19.2168 orf19.2177 orf19.2186 orf19.2202 orf19.2231 orf19.2244 orf19.2246 orf19.2247 orf19.2249 orf19.2256 orf19.2260 orf19.2262 orf19.2263 orf19.2282 orf19.2284 orf19.2285 orf19.2310 orf19.2312 orf19.2317 orf19.2318.1 orf19.2319 orf19.2328 orf19.2346 orf19.2350 orf19.2362 orf19.2366 orf19.2367 orf19.2386 orf19.2389 orf19.2392 orf19.2397 orf19.2399 orf19.2436 orf19.2445 orf19.2446 orf19.2457 orf19.246 orf19.2468 orf19.247 orf19.25 orf19.2517 orf19.2518 orf19.2529 orf19.254 orf19.2542 orf19.2544 orf19.2549 orf19.2552 orf19.2563 orf19.2591 orf19.2594 orf19.2607 orf19.2612 orf19.263.1 orf19.2631 orf19.2638 orf19.2639.1 orf19.2642 orf19.265 orf19.2663 orf19.267 orf19.2670 orf19.2675 orf19.27 orf19.270 orf19.2701 orf19.2708 orf19.2711 orf19.2721 orf19.2724 orf19.2726 orf19.2728 orf19.2731 orf19.2733 orf19.2737 orf19.2742 orf19.2749 orf19.276 orf19.2763 orf19.2769 orf19.2770 orf19.2778 orf19.2779 orf19.278 orf19.279 orf19.2795 orf19.2798 orf19.281 orf19.2812 orf19.2820 orf19.2835 orf19.2836 orf19.2847 orf19.2850 orf19.2852 orf19.2866 orf19.2870 orf19.2875 orf19.2887 orf19.2893 orf19.2908 orf19.2915 orf19.2917 orf19.2925 orf19.2930 orf19.2959.1 orf19.2962 orf19.2964 orf19.2965.1 orf19.2982 orf19.2996 orf19.3003 orf19.302 orf19.3022 orf19.3031 orf19.3042 orf19.3048 orf19.3051 orf19.3053 orf19.3055 orf19.3058 orf19.3060 orf19.3062 orf19.307 orf19.3073 orf19.3086 orf19.31 orf19.310 orf19.3100 orf19.3105 orf19.3107 orf19.3129 orf19.3148 orf19.3156 orf19.3161 orf19.3166 orf19.3178 orf19.3184 orf19.3185 orf19.3196 orf19.3204 orf19.3222 orf19.3232 orf19.3242 orf19.3248 orf19.3267 orf19.3272 orf19.3277 orf19.3280 orf19.3295 orf19.3296 orf19.3303 orf19.3304 orf19.3309 orf19.332 orf19.3329 orf19.3332 orf19.3333 orf19.3338 orf19.3351 orf19.3352 orf19.3367 orf19.3368 orf19.337 orf19.3371 orf19.3373 orf19.3378 orf19.3395 orf19.3399 orf19.34 orf19.3404 orf19.3418 orf19.3428 orf19.3432 orf19.344 orf19.3447 orf19.345 orf19.3470 orf19.3473 orf19.3477 orf19.351 orf19.3512 orf19.3517 orf19.3544 orf19.3545 orf19.3547 orf19.3552 orf19.3559 orf19.3569 orf19.3604 orf19.3605 orf19.3610 orf19.3621 orf19.3625 orf19.3643 orf19.3649 orf19.3665 orf19.3678 orf19.371 orf19.3778 orf19.3782.2 orf19.3788 orf19.3793 orf19.3810 orf19.3815 orf19.3817 orf19.3820 orf19.3828 orf19.3831 orf19.3836 orf19.3848 orf19.3863 orf19.3869 orf19.3871 orf19.3872 orf19.3885 orf19.3888.2 orf19.3898 orf19.3906 orf19.3908 orf19.3910 orf19.3919 orf19.392 orf19.3925 orf19.3926 orf19.3928 orf19.3932 orf19.3940 orf19.3957 orf19.3970 orf19.3977 orf19.4011

orf19.4021 orf19.4029 orf19.4062 orf19.4070 orf19.4078 orf19.4080 orf19.4088 orf19.4090
orf19.4091 orf19.4101 orf19.4104 orf19.4106 orf19.4110 orf19.4115 orf19.4121 orf19.4125
orf19.4128 orf19.413 orf19.4170 orf19.4173 orf19.4174 orf19.4182 orf19.419 orf19.4195
orf19.4204 orf19.4206 orf19.4220 orf19.4227 orf19.4243 orf19.4244 orf19.4245 orf19.4246
orf19.4248 orf19.4250 orf19.4258 orf19.4283 orf19.4287 orf19.4294 orf19.4315 orf19.4325
orf19.4326 orf19.4368 orf19.4372 orf19.4376 orf19.4388 orf19.439 orf19.4391 orf19.4399
orf19.4407 orf19.4428 orf19.443 orf19.4437 orf19.4439 orf19.4445 orf19.4455 orf19.4468
orf19.4474 orf19.4476 orf19.4479 orf19.448 orf19.4492 orf19.4496 orf19.4498 orf19.450
orf19.4503 orf19.4504 orf19.4511 orf19.4515 orf19.4518 orf19.4520 orf19.4521 orf19.4529
orf19.4532 orf19.4537 orf19.4538 orf19.4539 orf19.4544 orf19.4557 orf19.4577 orf19.4581
orf19.4594 orf19.4601 orf19.4607 orf19.4612 orf19.4626 orf19.4642 orf19.4652 orf19.4653
orf19.4654 orf19.4659 orf19.4666 orf19.4668 orf19.4675 orf19.4706 orf19.4707 orf19.4713
orf19.4724 orf19.4726 orf19.4735 orf19.475 orf19.4750 orf19.4751 orf19.4756 orf19.478
orf19.4791 orf19.4793 orf19.4795 orf19.4796 orf19.4799 orf19.4801 orf19.4824 orf19.4825
orf19.4830 orf19.4835 orf19.4837 orf19.4844 orf19.485 orf19.4864 orf19.4873 orf19.4875
orf19.4888 orf19.4894 orf19.4895 orf19.4896 orf19.4897 orf19.4901 orf19.4905 orf19.4911
orf19.4913 orf19.4914 orf19.4922 orf19.4923 orf19.494 orf19.4942 orf19.4946 orf19.4954
orf19.4962 orf19.4963 orf19.4964 orf19.4966 orf19.4970 orf19.4976 orf19.499 orf19.4996
orf19.500 orf19.5006.1 orf19.501 orf19.5012 orf19.5017 orf19.5022 orf19.5030 orf19.5037
orf19.5043 orf19.5049 orf19.5066 orf19.5067 orf19.5090 orf19.510 orf19.5126 orf19.5134
orf19.5139 orf19.5140 orf19.5141 orf19.5143 orf19.516 orf19.5161 orf19.5169 orf19.5177
orf19.518 orf19.519 orf19.5195 orf19.520 orf19.5204 orf19.5205 orf19.5206 orf19.5210 orf19.5220
orf19.5250 orf19.5256 orf19.5259 orf19.5267 orf19.5277 orf19.5278 orf19.5282 orf19.5284
orf19.5287 orf19.5290 orf19.5339 orf19.5346 orf19.536 orf19.5364 orf19.5391 orf19.5406
orf19.5409 orf19.542.2 orf19.5422 orf19.5431 orf19.5433 orf19.5459 orf19.547 orf19.5474
orf19.5486 orf19.5490 orf19.5514 orf19.5518 orf19.552 orf19.5536 orf19.5541 orf19.555
orf19.5566 orf19.557 orf19.5572 orf19.5573 orf19.5587 orf19.5589 orf19.5608 orf19.5609
orf19.5611 orf19.5614 orf19.5619 orf19.5626 orf19.5627 orf19.5646 orf19.5648 orf19.5655
orf19.5665 orf19.5675 orf19.5676 orf19.5686 orf19.5689 orf19.5722 orf19.5734 orf19.5767
orf19.5772 orf19.5780 orf19.5783 orf19.5784 orf19.5799 orf19.580 orf19.5808 orf19.5812
orf19.5813 orf19.5814 orf19.5817 orf19.5825 orf19.5826 orf19.5828 orf19.5835 orf19.5838
orf19.5840 orf19.5841 orf19.5843 orf19.585 orf19.5854.1 orf19.5862 orf19.5863 orf19.5871
orf19.5879 orf19.588 orf19.5896 orf19.5915 orf19.592 orf19.5929 orf19.5952 orf19.5954
orf19.596.2 orf19.5965 orf19.597 orf19.5978 orf19.5980 orf19.5985 orf19.5989 orf19.6013
orf19.6016 orf19.6017 orf19.6023 orf19.6025 orf19.6027 orf19.6030 orf19.604 orf19.6048
orf19.605 orf19.6061 orf19.6077 orf19.6079 orf19.6080 orf19.6083 orf19.6090 orf19.6117
orf19.6135 orf19.6137 orf19.6147 orf19.6184 orf19.6189 orf19.6195 orf19.6200 orf19.6211
orf19.6219 orf19.6220 orf19.6227 orf19.6252 orf19.6255 orf19.6260 orf19.6263 orf19.6277
orf19.6282 orf19.6288 orf19.6306 orf19.6318 orf19.6325.1 orf19.6351 orf19.6353 orf19.6355
orf19.6356 orf19.6360 orf19.639 orf19.6393 orf19.6398 orf19.6403 orf19.6408 orf19.6418
orf19.6434 orf19.6435 orf19.6443 orf19.6448 orf19.6453 orf19.6461 orf19.6477 orf19.6484
orf19.6487 orf19.6501 orf19.6503 orf19.6509 orf19.6518 orf19.6528 orf19.6539 orf19.6557
orf19.6563 orf19.6578 orf19.6589 orf19.6592 orf19.6597 orf19.6599.1 orf19.6612 orf19.6623
orf19.6625 orf19.6628 orf19.6637 orf19.6667 orf19.6678 orf19.6679 orf19.6681 orf19.6694
orf19.670.2 orf19.6703 orf19.6709 orf19.6719 orf19.6720 orf19.6722 orf19.6723 orf19.6726
orf19.6730 orf19.6732 orf19.6740 orf19.6741 orf19.6752 orf19.6755 orf19.6756 orf19.6758
orf19.6766 orf19.6769 orf19.6783 orf19.6786 orf19.6789 orf19.6804 orf19.6818 orf19.6828
orf19.6830 orf19.6840 orf19.6843 orf19.6852 orf19.6855 orf19.6858 orf19.6862 orf19.6866
orf19.6868 orf19.687 orf19.6873.1 orf19.6879 orf19.6883 orf19.6886 orf19.6888 orf19.6893
orf19.6899 orf19.69 orf19.69.2 orf19.6905 orf19.692 orf19.6921 orf19.693 orf19.6941 orf19.6973
orf19.6982 orf19.6989 orf19.6990 orf19.6992 orf19.6995 orf19.6996 orf19.6997 orf19.7010
orf19.7011 orf19.7013 orf19.7020 orf19.7023 orf19.7029 orf19.7056 orf19.7057 orf19.7063
orf19.7067 orf19.7073 orf19.7074 orf19.7078 orf19.7085 orf19.7121 orf19.7131 orf19.715
orf19.716 orf19.7160 orf19.7166 orf19.7184 orf19.7192 orf19.7204 orf19.7210 orf19.7225
orf19.7227 orf19.7234 orf19.7243 orf19.7244 orf19.7250 orf19.7252 orf19.7254 orf19.7261
orf19.7271 orf19.7279 orf19.7283 orf19.7288 orf19.7290 orf19.7291 orf19.7305 orf19.7306
orf19.7311 orf19.7329 orf19.733 orf19.7341 orf19.7353 orf19.7361 orf19.7366 orf19.7385
orf19.7398 orf19.7422 orf19.7426 orf19.7455 orf19.7457 orf19.7460 orf19.7478 orf19.7482

orf19.7490 orf19.7494 orf19.7497 orf19.750 orf19.7501 orf19.7512 orf19.7522 orf19.7531 orf19.7539.1 orf19.7545 orf19.7548 orf19.7567 orf19.7576 orf19.7589 orf19.7596 orf19.7598 orf19.7602 orf19.7620 orf19.7627 orf19.763 orf19.764 orf19.7642 orf19.7643 orf19.7644 orf19.7645 orf19.7646 orf19.7662 orf19.7664 orf19.77.1 orf19.773 orf19.776 orf19.781 orf19.808 orf19.809 orf19.812 orf19.813 orf19.822 orf19.827 orf19.841 orf19.855 orf19.863 orf19.864 orf19.880 orf19.899 orf19.910 orf19.929 orf19.93 orf19.94 orf19.948 orf19.952 orf19.962 orf19.966 orf19.972 orf19.993 orf19.996 orf19.997 orf19.998, AAH1, AAP1, AAT21, AAT22, ABC1, ABP140, ABP2, ACB1, ACF2, ACH1, ACO2, ADE4, ADE5,7, ADE8, AGP2, ALA1, ALG2, ALG5, ALG7, ALK6, ALK8, ALO1, AMO2, AOX1, AOX2, APA2, APL2, APN1, APN2, ARC1, ARC15, ARC40, ARE2, ARG1, ARG11, ARG3, ARG8, ARL1, ARL3, ARO2, ARO3, ARO7, ARO9, ARP1, ARP3, ARP9, ARR3, ARX1, ASE1, ASH2, ASM3, ASR1, ASR3, ATC1, ATG9, ATP20, AUT7, BDF1, BEM2, BEM3, BET2, BMS1, BMT3BMT7 BRG1 BRN1 BUB2 BUB3 BUD16 BUD22 BUD6 BUL1 BUR2 CAN1 CAP1 CAR1 CAT2 CCR4 CDC5 CEK1 CFL2 CFL4 CFL5 CGT1 CHO1 CHS2 CHS3 CHS5 CHT1 CHT3 CHT4 CIC1 CIP1 CKA2 CLN3 CNS1 CNT COX6 COX7 COX9 CPA1 CPY1 CRD2 CRH11 CRL1 CRZ1 CSA1 CSE4 CSH1 CSI2 CST5 CTA1 CTF1 CTN1 CTN3 CUP1 CUP2 CYB2 CYS3 DAC1 DAL5 DAL7 DAL8 DAL9 DAP2 DBP2 DBP3 DBP7 DCK2 DDI1 DES1 DFG10 DIP2 DJP1 DLD1 DOA1 DPM1 DPP1 DRS1 DUR1,2 DUR4 EAF6 EAF7 EBP7 ECE1 ECM1 ECM18 ECM3 ECM331 ECM38 ECM39 ECM42 EHT1 ELF1 EMP46 ENG1 ERB1 ERF1 ERG10 ERG26 ERG4 ERO1 ESS1 EXM2 EXO1 EXO70 FAA21 FAA2-1 FAL1 FAV1 FCA1 FCY23 FCY24 FGR23 FGR27 FGR28 FGR34 FGR50 FGR51 FHL1 FIL1 FMO1 FMO2 FMP45 FOX3 FRE9 FRS2 FTH1 FUM11 GAD1 GAL1 GAL10 GAL102 GAL7 GAP2 GAT1 GCA1 GCA2 GCD1 GCN20 GCN5 GCV1 GCV2 GCY1 GDH2 GDH3 GEF2 GFA1 GLE1 GLG2 GLR1 GPI7 GRE2 GRE3 GRP2 GSC1 GSL1 GST3 GTT13 GUT2 GZF3 HAK1 HAP3 HAP31 HAP41 HAP43 HAP5 HAS1 HBR1 HBR2 HBR3 HCM1 HCR1 HDA1 HEM1 HET1 HEX1 HGC1 HGT12 HGT13 HGT18 HGT20 HIS3 HMX1 HNM3 HNM4 HNT2 HOL1 HRT2 HSP31 HST1 HST3 HWP1 HYR1 HYR3 ICL1 IDI1 IDP2 IFA14 IFA21 IFD6 IFE2 IFF5 IFF6 IFF9 IFG3 IFM1 IFM3 IFR1 IHD2 IME2 IML2 INN1 IPL1 ISC1 ISY1 JEN1 JIP5 KAR3 KEL1 KEX2 KGD1 KIN2 KIN3 KIP1 KRE1 KRE30 KRR1 LAP3 LCB4 LDG3 LEM3 LEU3 LIP2 LIP3 LIP4 LIP5 LIP7 LTE1 LYP1 MAK16 MAK21 MAK5 MAL2 MAL31 MAS2 MBF1 MCA1 MCT1 MDR1 MEP2 MET14 MET15 MET28 MET3 MID1 MIG1 MIH1 MLC1 MLS1 MNN4-4 MNN7 MNN9 MOB2 MOH1 MPP10 MRF1 MRPL27 MRPL6 MRPL8 MSH2 MSS116 MTR2 MTW1 MUP1 MVB12 NAG3 NAG6 NAR1 NCP1 NCR1 NGG1 NOG1 NOP15 NOP4 NOP5 NSA1 NUP NUP60 OCA1 OPT1 OPT4 OPT7 OPT9 ORC3 OSM2 OXR1 OYE22 OYE23 OYE32 PBR1 PCL2 PDC12 PDC2 PDK2 PDR17 PDS5 PDX1 PES1 PEX13 PEX4 PGA10 PGA13 PGA17 PGA23 PGA26 PGA27 PGA28 PGA31 PGA32 PGA41 PGA43 PGA45 PGA52 PGA6 PGA60 PGA7 PHHB PHO113 PHO114 PHO23 PHO89 PHR1 PIL1 PIR1 PLB5 PNG2 POL93 POP4 POT1 POT1-2 POX18 PPE1 PRB1 PRC2 PRC3 PRI2 PRM1 PRN2 PRO1 PRP22 PRP39 PRR2 PRT1 PRX1 PSA2 PST1 PST2 PSY2 PTC6 PTC7 PUP3 PUS7 PUT1 PUT2 PUT4 PWP2 QDR1 RAD51 RAD54 RAS2 RAT1 RBD1 RBF1 RBR1 RBT4 RBT7 RCE1 RCL1 RCN1 REG1 REI1 REX2 RFX1 RFX2 RGS2 RKI1 RLP24 RNR3 ROT1 RPA190 RPC40 RPC53 RPD31 RPF2 RPG1A RPL35 RPL7 RPN6 RRN11 RRN3 RRP15 RRP6 RRP8 RRP9 RRS1 RSR1 RTA2 RTF1 SAC7 SAP1 SAP10 SAP4 SAP5 SAP6 SAP7 SAP8 SAP9 SAP98 SAP99 SAS10 SAS2 SCP1 SCS7 SCW11 SDC1 SDS22 SEC15 SEC2 SEC4 SEC65 SEF1 SEN2 SES1 SET3 SFT2 SFU1 SGO1 SGS1 SHE9 SIK1 SIR2 SIT1 SIZ1 SKO1 SLD5 SMC2 SMC3 SMF12 SMF3 SNF2 SNF7 SNO1 SOD2 SOD6 SOF1 SPB1 SPC2 SPC98 SPL1 SPO11 SPP1 SPR3 SPT10 SRP54 SSN3 SSP96 SST2 SSY1 STB5 STE18 STE23 STI1 STP3 SUA72 SUI3 SUL2 SUP35 SUV3 SWC4 SWI6 SYS3 TAF19 TAF60 TBF1 TBP1 TEA1 TEF4 TERT TES15 TFB3 THI4 TIF11 TIF3 TIF35 TLG2 TOR1 TPM2 TPO4 TPS1 TPT1 TRM1 TRP99 TSC11 TSR1 TSR2 TUB4 UBA4 UBC8 UBP6 UGA1 UGA11 UGA2 UGA4 UGA6 UGP1 ULP2 ULP3 USO5 UTP4 UTP8 UTP9 UTR2 VAS1 VMA10 VPS17 VPS2 VPS20 VPS21 VPS35 VPS4 VTC3 WH11 WHI3 WOR1 WOR2 WRS1 WSC2 YAH1 YEA4 YHM1 YHM2 YKU80 YMC1 YMC2 YOX1 YPD1 YPT31 YPT72 YUH2 ZCF2 ZCF20 ZCF22 ZCF29 ZCF35 ZCF37 ZCF38 ZCF39 ZRT1 ZUO1

Annexe F5: List of differentially expressed genes in T2KO1 strain.

orf19.1005, orf19.1007, orf19.1029, orf19.1030, orf19.1039, orf19.1041, orf19.1045, orf19.1070, orf19.1087, orf19.1091, orf19.11, orf19.1106, orf19.1117, orf19.1160, orf19.1162.1, orf19.1167, orf19.117, orf19.1181, orf19.1183, orf19.1214, orf19.1250, orf19.1253, orf19.1267, orf19.1268,

orf19.1277, orf19.1281, orf19.1287, orf19.1296, orf19.1305, orf19.1308, orf19.1318, orf19.132, orf19.1323, orf19.1326, orf19.1335, orf19.1340, orf19.1344, orf19.1356, orf19.1365, orf19.1369, orf19.137, orf19.1381, orf19.1383, orf19.1387, orf19.1388, orf19.1389, orf19.1421, orf19.1424, orf19.1426, orf19.1430, orf19.1449, orf19.1473, orf19.1480, orf19.1484, orf19.1491, orf19.1495, orf19.150, orf19.1502, orf19.1504, orf19.1522, orf19.1531, orf19.1532, orf19.1539orf19.154 orf19.1544 orf19.1547 orf19.1548 orf19.1549 orf19.1557 orf19.1562 orf19.1574 orf19.1578 orf19.1582 orf19.1584 orf19.1606 orf19.1608 orf19.1609 orf19.1611 orf19.1637 orf19.1639 orf19.1642 orf19.1654 orf19.1664 orf19.1675 orf19.1682 orf19.1691 orf19.1697 orf19.1707 orf19.1708 orf19.1720 orf19.1734 orf19.1745 orf19.175 orf19.1757 orf19.1765 orf19.1769 orf19.1771 orf19.1772 orf19.1774 orf19.1776 orf19.1785 orf19.1789 orf19.1791 orf19.1794 orf19.1796 orf19.1802 orf19.1808 orf19.1830 orf19.1833 orf19.1834 orf19.1841 orf19.1852 orf19.1856 orf19.1857 orf19.1867 orf19.187 orf19.1871 orf19.1878 orf19.188 orf19.190 orf19.1917 orf19.194 orf19.1940 orf19.1943 orf19.1950 orf19.1952 orf19.1968 orf19.1968.1 orf19.1999 orf19.200 orf19.2000 orf19.2018.1 orf19.2030 orf19.2041 orf19.2045 orf19.2048 orf19.2049 orf19.2050 orf19.2057 orf19.206 orf19.2064 orf19.2066.1 orf19.2067 orf19.2076 orf19.2090 orf19.2091 orf19.2114 orf19.2124 orf19.2125 orf19.215 orf19.2175 orf19.2177 orf19.2186 orf19.2202 orf19.2208 orf19.2227 orf19.2244 orf19.2247 orf19.2260 orf19.2262 orf19.2278 orf19.2284 orf19.2285 orf19.229 orf19.2301 orf19.2310 orf19.2314 orf19.2319 orf19.2320 orf19.2346 orf19.2366 orf19.2368 orf19.2376 orf19.2386 orf19.239 orf19.2392 orf19.2394 orf19.2397 orf19.2399 orf19.2400 orf19.2408 orf19.2414 orf19.2445 orf19.2449 orf19.246 orf19.2468 orf19.247 orf19.2499 orf19.2510 orf19.2515 orf19.2516 orf19.2517 orf19.2518 orf19.2529 orf19.2544 orf19.2552 orf19.2591 orf19.2594 orf19.2629 orf19.2631 orf19.2638 orf19.2639.1 orf19.264 orf19.2642 orf19.265 orf19.2650 orf19.2670 orf19.2680 orf19.2682 orf19.2684 orf19.2690 orf19.270 orf19.2701 orf19.2708 orf19.2720 orf19.2724 orf19.2728 orf19.2730 orf19.2731 orf19.2733 orf19.2734 orf19.2737 orf19.2739 orf19.2742 orf19.2749 orf19.2763 orf19.2769 orf19.2770 orf19.278 orf19.2781 orf19.279 orf19.2795 orf19.2798 orf19.2812 orf19.2818 orf19.2820 orf19.2825 orf19.2827 orf19.284 orf19.2846 orf19.2847 orf19.285 orf19.2850 orf19.286 orf19.2866 orf19.2867 orf19.287 orf19.2870 orf19.2875 orf19.2887 orf19.2899 orf19.2908 orf19.2915 orf19.2917 orf19.2920 orf19.2921 orf19.2938 orf19.2939 orf19.2940 orf19.2949 orf19.295 orf19.2957 orf19.2959.1 orf19.2965 orf19.2965.1 orf19.2986 orf19.2996 orf19.3009 orf19.3021 orf19.3029 orf19.3042 orf19.3049 orf19.3051 orf19.3053 orf19.31 orf19.310 orf19.3100 orf19.3105 orf19.3140.1 orf19.3156 orf19.3158 orf19.3161 orf19.3166 orf19.3178 orf19.318 orf19.3184 orf19.319 orf19.320 orf19.3205 orf19.3215 orf19.3216 orf19.3219 orf19.3222 orf19.3228 orf19.3237 orf19.3241 orf19.3245 orf19.3248 orf19.3250 orf19.3254 orf19.326 orf19.3267 orf19.3277 orf19.3295 orf19.3296 orf19.3303 orf19.3309 orf19.331 orf19.3310 orf19.3317 orf19.332 orf19.3323 orf19.3338 orf19.3353 orf19.3366.1 orf19.3368 orf19.3371 orf19.3373 orf19.3378 orf19.3399 orf19.34 orf19.3404 orf19.3407 orf19.341 orf19.3418 orf19.3425 orf19.3444 orf19.3447 orf19.3448 orf19.3449 orf19.3456 orf19.3470 orf19.3473 orf19.3475 orf19.3477 orf19.3483 orf19.351 orf19.3512 orf19.3545 orf19.3547 orf19.3569 orf19.3572 orf19.3601 orf19.3604 orf19.3607 orf19.3610 orf19.3613 orf19.3615 orf19.3617 orf19.3621 orf19.3624 orf19.3625 orf19.3643 orf19.3667 orf19.3678 orf19.3704 orf19.3714 orf19.372 orf19.3724 orf19.3737 orf19.3763 orf19.3767 orf19.3788 orf19.3793 orf19.3813 orf19.3815 orf19.3820 orf19.3828 orf19.3831 orf19.3843 orf19.3848 orf19.3858 orf19.3863 orf19.3869 orf19.3871 orf19.3902 orf19.3905 orf19.3910 orf19.3919 orf19.3920 orf19.3924 orf19.3928 orf19.3929 orf19.3932 orf19.3942 orf19.3946 orf19.3956 orf19.3957 orf19.399 orf19.3990 orf19.3999 orf19.4007 orf19.4011 orf19.4017 orf19.4043 orf19.4068 orf19.4070 orf19.4078 orf19.4080 orf19.4090 orf19.4091 orf19.4104 orf19.4105 orf19.4106 orf19.411 orf19.4115 orf19.4116 orf19.4121 orf19.4123 orf19.4128 orf19.413 orf19.4143 orf19.4159 orf19.416 orf19.4163 orf19.417 orf19.4170 orf19.4173 orf19.4176 orf19.419 orf19.4194 orf19.4203 orf19.4204 orf19.4220 orf19.4227 orf19.4229 orf19.4230 orf19.4243 orf19.4244 orf19.4246 orf19.4250 orf19.4258 orf19.4269 orf19.4281 orf19.4287 orf19.4293 orf19.4294 orf19.4312 orf19.4315 orf19.4316 orf19.432 orf19.4324 orf19.4326 orf19.4330 orf19.4337 orf19.4340 orf19.4342 orf19.4348 orf19.4349 orf19.4362 orf19.4365 orf19.4391 orf19.4400 orf19.4428 orf19.443 orf19.4437 orf19.4439 orf19.4449 orf19.4455 orf19.4459 orf19.4465 orf19.4468 orf19.4476 orf19.4479 orf19.4488 orf19.449 orf19.4492 orf19.4498 orf19.450 orf19.4503 orf19.4504 orf19.4511 orf19.4518 orf19.4530.1 orf19.4544 orf19.4557 orf19.4577 orf19.4594 orf19.4607 orf19.4614 orf19.4615 orf19.4617 orf19.4620 orf19.4621 orf19.4622 orf19.4623 orf19.4626 orf19.4627 orf19.4629 orf19.4642 orf19.4652 orf19.4653 orf19.4658

orf19.4659 orf19.4668 orf19.4675 orf19.4691 orf19.4706 orf19.4711 orf19.4713 orf19.4728
orf19.4730 orf19.4731 orf19.4735 orf19.474 orf19.4740 orf19.4751 orf19.4764 orf19.4768
orf19.4791 orf19.4792 orf19.4793 orf19.4795 orf19.4799 orf19.4812 orf19.4824 orf19.4825
orf19.4835 orf19.4837 orf19.4850 orf19.4875 orf19.4880 orf19.4883 orf19.4888 orf19.4889
orf19.4894 orf19.4896 orf19.4897 orf19.4898 orf19.4900 orf19.4901 orf19.4913 orf19.4919
orf19.4946 orf19.4947 orf19.496 orf19.4963 orf19.4966 orf19.4970 orf19.4988 orf19.499
orf19.4996 orf19.50 orf19.500 orf19.5003 orf19.5008.1 orf19.5009 orf19.501 orf19.5019 orf19.5030
orf19.5037 orf19.5038 orf19.5043 orf19.5049 orf19.5051 orf19.5066 orf19.5067 orf19.5078
orf19.5134 orf19.5139 orf19.5140 orf19.5141 orf19.515 orf19.5165 orf19.5169 orf19.519
orf19.5194.1 orf19.5195 orf19.520 orf19.5204 orf19.5205 orf19.5221 orf19.5225 orf19.5239
orf19.5247 orf19.5250 orf19.5266 orf19.5275 orf19.5276 orf19.5277 orf19.5284 orf19.5287
orf19.5290 orf19.5297 orf19.530 orf19.5312 orf19.5342 orf19.536 orf19.5365 orf19.5394.1
orf19.5406 orf19.541 orf19.5411 orf19.5413 orf19.5428 orf19.5431 orf19.5455 orf19.547
orf19.5483 orf19.5488 orf19.5490 orf19.5510 orf19.5514 orf19.5516 orf19.552 orf19.5532
orf19.5536 orf19.5541 orf19.5552 orf19.5565 orf19.557 orf19.5572 orf19.5573 orf19.5575
orf19.5587 orf19.5605 orf19.5608 orf19.5609 orf19.5614 orf19.5616 orf19.5617 orf19.5618
orf19.5619 orf19.5627 orf19.5628 orf19.5646 orf19.5655 orf19.5660 orf19.5676 orf19.5678
orf19.5683 orf19.5704 orf19.5722 orf19.5747 orf19.5772 orf19.578 orf19.5782 orf19.5784
orf19.5785 orf19.5799 orf19.5802 orf19.5808 orf19.5809 orf19.5812 orf19.5817 orf19.5826
orf19.5831 orf19.5833 orf19.5841 orf19.5842.1 orf19.5843 orf19.5854.1 orf19.5871 orf19.5876
orf19.588 orf19.5897 orf19.5905 orf19.5915 orf19.592 orf19.5925 orf19.5934 orf19.5942
orf19.5954 orf19.5961 orf19.597 orf19.5978 orf19.5980 orf19.5984 orf19.5985 orf19.6012
orf19.6013 orf19.6017 orf19.6027 orf19.6043 orf19.6048 orf19.6061 orf19.6066 orf19.6079
orf19.6114 orf19.6117 orf19.6143 orf19.6147 orf19.6152 orf19.6184 orf19.6189 orf19.6195
orf19.6196 orf19.6211 orf19.6230 orf19.6236 orf19.6252 orf19.6255 orf19.6260 orf19.6264.4
orf19.6271 orf19.6272 orf19.6275 orf19.6277 orf19.6282 orf19.6288 orf19.6297 orf19.6306
orf19.6310 orf19.6316.4 orf19.6325.1 orf19.6326 orf19.6346 orf19.6348 orf19.6350 orf19.6351
orf19.6353 orf19.6355 orf19.6357 orf19.6360 orf19.6374 orf19.639 orf19.6400 orf19.6403
orf19.6408 orf19.6416 orf19.643 orf19.6443 orf19.6453 orf19.6456 orf19.6460 orf19.6484
orf19.649 orf19.6503 orf19.6508 orf19.6509 orf19.6518 orf19.652 orf19.6522 orf19.6528
orf19.6530 orf19.6537 orf19.6550 orf19.6557 orf19.6578 orf19.6583 orf19.6587 orf19.6589
orf19.6597 orf19.6601 orf19.6612 orf19.6623 orf19.6625 orf19.6628 orf19.6637 orf19.6662
orf19.6678 orf19.668 orf19.6681 orf19.6688 orf19.6694 orf19.6698 orf19.670.2 orf19.6703
orf19.6704 orf19.6705 orf19.6713 orf19.6720 orf19.6730 orf19.6731 orf19.6740 orf19.6741
orf19.6742 orf19.6753 orf19.6754 orf19.6758 orf19.6766 orf19.6786 orf19.6788 orf19.6789
orf19.6795 orf19.6797 orf19.6804 orf19.6805 orf19.6818 orf19.6822 orf19.6828 orf19.6838
orf19.6840 orf19.6843 orf19.6845 orf19.6855 orf19.6861 orf19.6862 orf19.6868 orf19.6869
orf19.687 orf19.6873.1 orf19.6880 orf19.6886 orf19.6888 orf19.6893 orf19.6898 orf19.69.2
orf19.6903 orf19.6912 orf19.6921 orf19.6952 orf19.6966 orf19.6973 orf19.6982 orf19.6986
orf19.6990 orf19.6992 orf19.6995 orf19.6996 orf19.6997 orf19.7009 orf19.7010 orf19.7011
orf19.7023 orf19.7028 orf19.7029 orf19.7038 orf19.7056 orf19.7057 orf19.7061 orf19.7067
orf19.7077 orf19.7078 orf19.7086 orf19.7098 orf19.7111 orf19.7121 orf19.7130 orf19.7131
orf19.715 orf19.7151 orf19.7184 orf19.7192 orf19.7197 orf19.7204 orf19.721 orf19.7210
orf19.7225 orf19.7227 orf19.7234 orf19.7244 orf19.725 orf19.7250 orf19.7267 orf19.7279
orf19.7283 orf19.7288 orf19.729 orf19.7305 orf19.7306 orf19.7310 orf19.732 orf19.733 orf19.7344
orf19.7347 orf19.7353 orf19.7361 orf19.7365 orf19.7366 orf19.7385 orf19.7386 orf19.7398
orf19.7410 orf19.7422 orf19.7441 orf19.7449 orf19.7450 orf19.7457 orf19.7473 orf19.7482
orf19.7488 orf19.7495 orf19.75 orf19.7501 orf19.7507 orf19.7512 orf19.7527 orf19.7531
orf19.7539.1 orf19.7545 orf19.7546 orf19.7552 orf19.7553 orf19.7554 orf19.7566 orf19.7567
orf19.7576 orf19.7589 orf19.7593 orf19.7596 orf19.7620 orf19.7627 orf19.7643 orf19.7646
orf19.773 orf19.781 orf19.789.1 orf19.792 orf19.8 orf19.808 orf19.809 orf19.810 orf19.812
orf19.813 orf19.822 orf19.828 orf19.839 orf19.841 orf19.855 orf19.879 orf19.899 orf19.910
orf19.921 orf19.928 orf19.929 orf19.93 orf19.949 orf19.952 orf19.962 orf19.996 AAH1 AAP1
AAT21 AAT22 ABC1 ABP140 ACE2 ACO2 ADE8 AFG1 AFT2 AGP3 AIP2 ALG7 ALK2 ALP1
ALS1 ALS6 AMO2 AOX1 AOX2 APL5 APN2 AQY1 ARE2 ARG1 ARG11 ARG3 ARG4 ARG5,6
ARG8 ARG83 ARL1 ARL3 ARO10 ARP1 ARP3 ARP4 ARP8 ARP9 ARX1 ASE1 ASH2 ASM3
ASR1 ASR3 ATC1 ATG9 AUR1 AXL1 AYR2 BAT21 BBC1 BEM2 BGL2 BLM3 BMS1 BMT5 BMT8
BRG1 BRN1 BUB1 BUB3 BUD2 BUD22 BUL1 BUR2 BZZ1 CAN1 CAN2 CAP1 CAR1 CAS1 CBF1

CCC1 CDC14 CDC27 CDC3 CDC68 CDR1 CEK2 CFL2 CGR1 CGT1 CHA1 CHK1 CHO1 CHO2 CHR1 CHS1 CHT1 CIC1 CIP1 CLB4 CLG1 CLN3 CNT COX5 COX9 CPA1 CPA2 CPY1 CRD2 CRH11 CRH12 CRL1 CSA1 CSH1 CSO99 CSP37 CST5 CTA2 CTF18 CTN1 CTN3 CTR2 CUP1 CUP9 CYB2 CYC1 CYS3 CZF1 DAC1 DAG7 DAL5 DAL52 DAL7 DAL8 DAL9 DAO1 DAP1 DBP2 DBP3 DBP7 DDI1 DDR48 DFG10 DFI1 DFR1 DIP5 DJP1 DLH1 DOA4 DRS1 DSL1 DTD2 DUN1 DUR1,2 DUR4 DUR7 EAF3 EAF7 EBP7 ECM1 ECM15 ECM17 ECM18 ECM21 ECM3 ECM331 ECM42 ELF1 ELP3 END3 ENP1 ENP2 ERB1 ERG6 ERO1 ESP1 EXG2 EXO70 FAA21 FAB1 FAL1 FAV1 FCR1 FCY2 FCY24 FDH1 FET3 FET34 FGR27 FGR28 FGR32 FGR50 FGR51 FIL1 FLU1 FMO1 FMP45 FOL1 FPG1 FRE7 FRP2 FTH1 FTH2 FUS1 GAL10 GAL102 GAP2 GAT1 GCA1 GCA2 GCD7 GCF1 GCN5 GCV1 GDB1 GDH2 GEA2 GFA1 GIN1 GLE1 GLG2 GLO1 GLR1 GNP3 GPH1 GRE2 GRP2 GSC1 GSG1 GSL1 GTT12 GYP1 GYP7 HAK1 HAP2 HAP41 HBR3 HCA4 HDA1 HEM1 HEM15 HEX1 HGT12 HGT13 HGT16 HGT18 HGT5 HIS3 HIS4 HMI1 HNM1 HNM3 HNM4 HNT1 HOC1 HOL1 HOS1 HRK1 HRT2 HSP31 HST1 HST6 HXT5 ICL1 IFA21 IFC1 IFD6 IFE1 IFE2 IFF4 IFF6 IFF9 IFG3 IFH1 IFM3 IFR1 IFU5 IHD1 IML2 INO1 IPL1 ISA1 ISC1 IST2 ISW2 ISY1 ITR1 JEN1 KAR3 KAR9 KEM1 KEX2 KGD1 KGD2 KIN2 KIN3 KIP1 KIP2 KIP4 KIS1 KNS1 KRE30 KRR1 LAG1 LAP3 LDG3 LEU3 LEU4 LEU42 LIP2 LIP3 LIP4 LIP5 LIP7 LTE1 LYP1 LYS2 MAC1 MAD2 MAK16 MAK21 MAK32 MAK5 MAL31 MBF1 MCI4 MCT1 MDN1 MEP2 MET14 MET15 MET3 MEU1 MHP1 MID1 MIH1 MIT1 MLC1 MNN1 MNN4 MNN7 MNN9 MNR2 MOH1 MRE11 MRF1 MRPL33 MSH2 MSN5 MSS116 MTW1 MUC1 MUM2 MVB12 MXR1 MYO1 NAB3 NAG3 NAG6 NCR1 NHP2 NIP1 NIP100 NIT3 NOC2 NOG1 NOP14 NOP15 NOP4 NOP5 NPL6 NRG1 NRP1 NSA1 NUF2 NUP49 NUP84 NUP85 OPT1 OPT4 OPT9 OSM2 OYE2 OYE22 OYE23 OYE32 PCK1 PCL1 PCL2 PCL7 PCT1 PDC12 PDE1 PDE2 PDK2 PDR17 PDS5 PEP7 PES1 PEX13 PEX14 PEX4 PGA10 PGA13 PGA23 PGA26 PGA32 PGA41 PGA43 PGA60 PGA7 PHHB PHO113 PHO114 PHO23 PHO84 PHO89 PHR1 PIR1 PLB1 PLB4.5 PLB5 PMS1 PNG2 POL2 POL3 POL32 POL5 POL93 POP3 POP4 PPT1 PPZ1 PR26 PRB1 PRC2 PRE1 PRE5 PRI2 PRM1 PRN2 PRN3 PRO3 PRP39 PRR2 PRT1 PSA2 PST1 PST2 PSY2 PTC4 PTC5 PTC7 PTH2 PTP1 PUF3 PUP3 PUS7 PUT1 PUT2 PUT4 PWP2 PXP2 QCR8 RAD16 RAD54 RAD7 RAD9 RAM1 RAS2 RBF1 RBR1 RBT4 RBT7 RCE1 RCN1 RCY1 REG1 REI1 REV3 REX2 RFX2 RGA2 RGS2 RIM2 RLM1 RLP24 RMT2 RNR3 RPA190 RPB4 RPC19 RPC31 RPC40 RPC53 RPD31 RPF2 RPG1A RPL35 RPL7 RPN10 RPN4 RPS27 RRN3 RRP6 RRP8 RRS1 RSR1 RTA3 RTF1 RVS162 SAC7 SAP1 SAP10 SAP2 SAP5 SAP6 SAP8 SAP9 SAP99 SAS10 SAS2 SAS3 SCP1 SCS7 SDA1 SDH1 SDH2 SDH4 SDS24 SEC1 SEC2 SEN2 SET2 SET3 SFC1 SFL1 SFT2 SFU1 SGE1 SGO1 SIK1 SIR2 SIT1 SIZ1 SKN2 SKO1 SLA2 SLN1 SMC2 SMC3 SMC4 SMC5 SMF12 SMM1 SNF2 SNG3 SNO1 SNU114 SNX4 SOD1 SOD2 SOD6 SOU1 SPB1 SPB4 SPL1 SPO75 SPT10 SPT6 SRP40 SSN3 SSP96 SST2 STE18 STI1 SUA72 SUP35 SUR2 SVF1 SWC4 SYG1 SYS3 TAF145 TAF19 TAF60 TBF1 TEA1 TERT TES15 TFB3 TFS1 THI4 TIF11 TIF3 TIM9 TLG2 TLO1 TLO16 TOP1 TOP2 TPK1 TPM2 TPS1 TPS2 TRM1 TRP99 TSR2 TTR1 TUB4 UBA4 UBP1 UGA1 UGA11 UGA2 UGA4 UGA6 UGP1 ULP3 URA1 USO5 USO6 UTP4 UTR2 VAN1 VMA22 VPS20 VPS21 VTC3 WH11 WOR1 WOR2 WRS1 WSC1 WSC2 XOG1 XUT1 YCG1 YMC1 YMC2 YNK1 YOR1 YPD1 YVH1 ZCF11 ZCF2 ZCF20 ZCF22 ZCF25 ZCF26 ZCF31 ZCF32 ZCF37 ZCF39 ZCF4 ZCF9 ZDS1 ZRT2 ZUO1

Annexe F6: List of differentially expressed genes in T2KO2 strain.

orf19.1007, orf19.1028, orf19.1041, orf19.1043, orf19.1075, orf19.1082.1, orf19.1085, orf19.1087, orf19.109, orf19.1095, orf19.1107, orf19.1113, orf19.1121, orf19.1122, orf19.1124, orf19.1124.2, orf19.1143, orf19.1144, orf19.1167, orf19.1171, orf19.1182, orf19.1200, orf19.1207, orf19.1210, orf19.1257, orf19.1268, orf19.1270, orf19.1272, orf19.1286, orf19.1296, orf19.1304, orf19.1306, orf19.1326, orf19.1334, orf19.1343, orf19.1344, orf19.1356, orf19.1363, orf19.1371, orf19.1383, orf19.1386, orf19.1394, orf19.1427, orf19.1433, orf19.1449, orf19.1477, orf19.1491, orf19.150, orf19.1505, orf19.1514, orf19.1522, orf19.1531, orf19.1535, orf19.1549, orf19.1562, orf19.1632, orf19.1639, orf19.164, orf19.1641, orf19.1646, orf19.1656, orf19.1664, orf19.1668orf19.1682 orf19.1684 orf19.1709 orf19.1717 orf19.1723 orf19.1728 orf19.1735 orf19.1737 orf19.1748 orf19.175 orf19.1799 orf19.1801 orf19.1808 orf19.1827 orf19.1834 orf19.1840 orf19.1841 orf19.1852 orf19.1857 orf19.1867 orf19.188 orf19.1881 orf19.1887 orf19.1888 orf19.190 orf19.1910 orf19.1938 orf19.1940 orf19.1950 orf19.1952 orf19.1959 orf19.1964 orf19.1968 orf19.1970 orf19.1972 orf19.1975 orf19.1981 orf19.1985 orf19.199 orf19.1991 orf19.1994

orf19.2006 orf19.2018.1 orf19.2035 orf19.2057 orf19.206 orf19.2067 orf19.2072 orf19.2076
orf19.2088 orf19.2091 orf19.2112 orf19.2114 orf19.2124 orf19.2128 orf19.2165 orf19.2166
orf19.2180 orf19.2213 orf19.2227 orf19.2247 orf19.2249 orf19.2262 orf19.2263 orf19.2269
orf19.2274 orf19.230 orf19.2301 orf19.2310 orf19.2312 orf19.2317 orf19.2325 orf19.2330
orf19.2366 orf19.2367 orf19.2376 orf19.2391 orf19.2397 orf19.2399 orf19.2408 orf19.2445
orf19.247 orf19.2487 orf19.2496 orf19.2510 orf19.252 orf19.254 orf19.2542 orf19.2574 orf19.258
orf19.2583.2 orf19.2590 orf19.2591 orf19.2610 orf19.2629 orf19.2638 orf19.2639.1 orf19.2663
orf19.2669 orf19.2684 orf19.2686 orf19.2697 orf19.270 orf19.2701 orf19.2733 orf19.2739
orf19.2749 orf19.2755 orf19.276 orf19.2763 orf19.2769 orf19.2770 orf19.2782 orf19.2784
orf19.279 orf19.2796 orf19.2798 orf19.2805 orf19.2812 orf19.2821 orf19.2822 orf19.2852
orf19.287 orf19.2887 orf19.2890 orf19.2892 orf19.2893 orf19.291 orf19.2916 orf19.2920
orf19.2930 orf19.2953 orf19.2959.1 orf19.2963 orf19.2964 orf19.2978 orf19.2985 orf19.3007.2
orf19.3016 orf19.3029 orf19.3043 orf19.3045 orf19.3049 orf19.3061 orf19.3062 orf19.3088
orf19.3089 orf19.310 orf19.3105 orf19.3113 orf19.3124 orf19.3139 orf19.3140.1 orf19.3148
orf19.315 orf19.3156 orf19.3163 orf19.3166 orf19.3167 orf19.3169 orf19.3170 orf19.3173
orf19.3175 orf19.3177 orf19.3178 orf19.3184 orf19.3185 orf19.3213 orf19.3214 orf19.3222
orf19.3237 orf19.3248 orf19.3283 orf19.3290 orf19.3295 orf19.3306 orf19.3317 orf19.3318
orf19.3338 orf19.3352 orf19.3366.1 orf19.338 orf19.3399 orf19.34 orf19.341 orf19.3418 orf19.3432
orf19.3442 orf19.345 orf19.3473 orf19.3512 orf19.354 orf19.3544 orf19.3573 orf19.3578
orf19.3604 orf19.3610 orf19.3613 orf19.3621 orf19.3625 orf19.3643 orf19.3644 orf19.3663.1
orf19.3684 orf19.3689 orf19.3763 orf19.3767 orf19.3774.1 orf19.3784 orf19.3793 orf19.3799
orf19.3848 orf19.3863 orf19.3869 orf19.3872 orf19.3897 orf19.3898 orf19.3901 orf19.3908
orf19.3915 orf19.3920 orf19.3921 orf19.3924 orf19.3926 orf19.3929 orf19.3932 orf19.3936
orf19.3937 orf19.3939 orf19.3942 orf19.3947 orf19.3952 orf19.3954 orf19.3956 orf19.3957
orf19.3963 orf19.398 orf19.3980 orf19.3991 orf19.3996 orf19.4005 orf19.4006 orf19.4016
orf19.4017 orf19.4018 orf19.4019 orf19.4021 orf19.4030 orf19.4031 orf19.4070 orf19.4080
orf19.4086 orf19.4091 orf19.4104 orf19.4121 orf19.4134 orf19.4144 orf19.4148 orf19.4171
orf19.4182 orf19.4194 orf19.4203 orf19.4210 orf19.4227 orf19.4228 orf19.4230 orf19.4240
orf19.4241 orf19.4245 orf19.4248 orf19.4250 orf19.4252 orf19.4281 orf19.4283 orf19.4292
orf19.4293 orf19.4295 orf19.4307 orf19.4316 orf19.433 orf19.4332 orf19.4340 orf19.4346
orf19.4347 orf19.4373 orf19.4398 orf19.4437 orf19.4459 orf19.4465 orf19.4476 orf19.4478
orf19.4488 orf19.4498 orf19.4504 orf19.4507 orf19.4523 orf19.4538 orf19.4539 orf19.4557
orf19.4567 orf19.4569 orf19.4571 orf19.4600.1 orf19.4610 orf19.4620 orf19.4621 orf19.4629
orf19.4634 orf19.4639 orf19.4642 orf19.4668 orf19.467 orf19.4675 orf19.4680 orf19.4690
orf19.4691 orf19.4706 orf19.4723 orf19.4726 orf19.4731 orf19.4750 orf19.4763 orf19.478
orf19.4783 orf19.479.2 orf19.4792 orf19.4805 orf19.4861 orf19.4864 orf19.4865 orf19.4873
orf19.4880 orf19.4883 orf19.4894 orf19.4904 orf19.4905 orf19.4906 orf19.4912 orf19.4913
orf19.4931 orf19.4946 orf19.4963 orf19.4966 orf19.4976 orf19.4988 orf19.499 orf19.5017
orf19.5026 orf19.5037 orf19.5038 orf19.5041 orf19.5042 orf19.5043 orf19.5052 orf19.5070
orf19.5125 orf19.5141 orf19.5160 orf19.519 orf19.5195 orf19.5258 orf19.5259 orf19.5262
orf19.5266 orf19.5267 orf19.5270 orf19.5284 orf19.5287 orf19.5295 orf19.5300 orf19.5340
orf19.5345 orf19.5350 orf19.5364 orf19.5368 orf19.5376 orf19.5393 orf19.540 orf19.5406
orf19.5418 orf19.5450 orf19.5490 orf19.55 orf19.5518 orf19.552 orf19.5525 orf19.5532 orf19.5535
orf19.5565 orf19.5575 orf19.5579 orf19.5587 orf19.5589 orf19.5611 orf19.5614 orf19.5616
orf19.5619 orf19.5620 orf19.5655 orf19.5675 orf19.5680 orf19.5681 orf19.5682 orf19.5684
orf19.5689 orf19.5692 orf19.5693 orf19.5698 orf19.5722 orf19.5763 orf19.5772 orf19.5780
orf19.5784 orf19.5799 orf19.580 orf19.581 orf19.5831 orf19.5838 orf19.5840 orf19.5863
orf19.5865 orf19.5876 orf19.588 orf19.5896 orf19.5915 orf19.5952 orf19.5954 orf19.597
orf19.5973 orf19.5984 orf19.5985 orf19.6008 orf19.6012 orf19.6017 orf19.6027 orf19.605
orf19.6066 orf19.6076 orf19.6079 orf19.6083 orf19.609 orf19.6102 orf19.6114 orf19.6132
orf19.6138 orf19.6143 orf19.6160 orf19.6171 orf19.6184 orf19.6187 orf19.6209 orf19.6246
orf19.6248 orf19.6255 orf19.6260 orf19.6277 orf19.6305 orf19.6310 orf19.6311 orf19.6316
orf19.6357 orf19.6360 orf19.6366 orf19.6382 orf19.6407 orf19.6424 orf19.6435 orf19.6443
orf19.6456 orf19.6457 orf19.647.3 orf19.6484 orf19.6524 orf19.6553 orf19.6555 orf19.6566
orf19.6580 orf19.6583 orf19.6591 orf19.6592 orf19.6605 orf19.6606 orf19.6628 orf19.6654
orf19.6662 orf19.6665 orf19.6667 orf19.6668 orf19.6676 orf19.6678 orf19.668 orf19.6681
orf19.6709 orf19.6713 orf19.6720 orf19.6722 orf19.6731 orf19.6740 orf19.6747 orf19.6753
orf19.6754 orf19.6755 orf19.6756 orf19.6759 orf19.6769 orf19.6786 orf19.6789 orf19.6793

orf19.6807 orf19.6810 orf19.6811 orf19.6822 orf19.6828 orf19.6838 orf19.6861 orf19.6864 orf19.6874 orf19.6880 orf19.6886 orf19.6899 orf19.6912 orf19.692 orf19.6921 orf19.6950 orf19.6966 orf19.6968 orf19.6981 orf19.6982 orf19.6986 orf19.6989 orf19.6992 orf19.6996 orf19.7013 orf19.7023 orf19.7032 orf19.7038 orf19.7042 orf19.7056 orf19.7069 orf19.7070 orf19.7077 orf19.7082 orf19.7098 orf19.7100 orf19.7110 orf19.7111 orf19.7119 orf19.7130 orf19.7131 orf19.7149 orf19.715 orf19.7152 orf19.7166 orf19.7184 orf19.7192 orf19.7193 orf19.7199 orf19.7200 orf19.7214 orf19.7215 orf19.7225 orf19.725 orf19.7269 orf19.7279 orf19.7285 orf19.7288 orf19.7291 orf19.7296 orf19.7304 orf19.7310 orf19.7330 orf19.7336 orf19.7337 orf19.7344 orf19.7353 orf19.7368 orf19.7398 orf19.7403 orf19.7410 orf19.7420 orf19.7422 orf19.7442 orf19.7452 orf19.7468 orf19.7488 orf19.7495 orf19.750 orf19.7546 orf19.7566 orf19.757 orf19.7588 orf19.7601 orf19.7617 orf19.7618 orf19.7645 orf19.7646 orf19.7648 orf19.7662 orf19.792 orf19.804 orf19.808 orf19.855 orf19.86 orf19.873 orf19.891 orf19.9 orf19.91 orf19.929 orf19.93 orf19.932 orf19.933 orf19.94 orf19.952 orf19.961.2 orf19.966 orf19.967 orf19.969 orf19.970 orf19.980 orf19.993 AAH1 AAP1 AAT1 AAT22 ABP140 ACB1 ACC1 ACF2 ACS1 ADE4 ADE5,7 ADH3 AGP3 AIP2 ALD6 ALG2 ALG5 ALG6 ALK8 ALP1 ALR1 ANT1 APN1 AQY1 ARF3 ARG1 ARG11 ARG3 ARG4 ARG5,6 ARG8 ARG83 ARO3 ARP1 ARP9 ARR3 ASE1 ASR1 ASR2 ATC1 ATG9 AXL1 AYR1 AYR2 BAT21 BCK1 BDF1 BEM3 BET2 BGL2 BMT3 BMT6 BUD23 BUD6 BUL1 CAC2 CAG1 CAK1 CAM1 CAN2 CAR1 CAR2 CAT1 CAT2 CBP1 CCC1 CCR4 CCT2 CCT3 CCT5 CCT7 CDC14 CDC20 CDC21 CDC24 CDC3 CDC5 CDC60 CDC68 CDG1 CDS1 CFL2 CGR1 CHA1 CHT3 CIP1 CIS2 CIT1 CKB1 CLB2 CNB1 COF1 COX5 COX6 COX9 CPA1 CPA2 CRD2 CRH11 CRL1 CSH1 CTA1 CTF18 CTF8 CTN1 CTP1 CTR1 CUP9 CYB2 CYS3 DAL1 DAL4 DAL7 DAO1 DBF2 DBP8 DCC1 DDI1 DDR48 DFI1 DIM1 DIP5 DLD1 DOG1 DPP1 DRG1 DUR1,2 DUR4 EAF3 EAF7 ECI1 ECM15 ECM17 ECM18 ECM3 ECM38 ECM42 EFH1 ELF1 END3 ERG3 ERG4 ERV46 EST1 EXG2 FAA21 FAR1 FAS2 FAT1 FAV1 FCA1 FCY21 FCY24 FESUR1 FET3 FET33 FET99 FGR22 FGR23 FGR27 FGR29 FIL1 FMO1 FMO2 FOX2 FOX3 FRE10 FRE7 FTR1 FUR1 FUR4 FUS1 GAC1 GAL1 GAL10 GAL102 GAL7 GAP1 GAP2 GAT1 GCA1 GCA2 GCD1 GCD11 GCD7 GCF1 GDH3 GLC3 GLE1 GLO1 GLR1 GLT1 GPR1 GPT1 GPX1 GRR1 GRX1 GST2 GST3 GTT12 GTT13 GUP1 GYP2 HAP2 HAP43 HAS1 HCH1 HCR1 HEM1 HEX1 HGH1 HGT1 HGT13 HGT16 HGT19 HGT2 HGT5 HGT6 HIS1 HIS3 HIS4 HIS5 HNM3 HNM4 HOC1 HOL1 HOS1 HRT2 HSL1 HST1 HTS1 HXT5 HYR1 HYS2 IDP2 IFC1 IFC3 IFD6 IFE1 IFG3 IHD1 ILV1 ILV3 IMP4 INO1 IQG1 ISA1 IST2 ISU1 ITR1 JEN1 JEN2 KAR9 KGD1 KIP1 KIP4 KSR1 KTI12 KTR4 LDG3 LEU42 LIG1 LIP5 LRO1 LTE1 LTV1 LYS1 LYS4 MAC1 MAD2 MAS2 MBF1 MCI4 MCM1 MCT1 MDM34 MDN1 MEC1 MEP2 MES1 MET14 MET16 MET3 MEU1 MIF2 MIT1 MKK2 MNN22 MNR2 MNS1 MPP10 MRP2 MRS2 MSK1 MTLA1 MTR2 MTW1 MUM2 MVB12 MXR1 NAB3 NAM7 NAT2 NCE103 NDH51 NIP7 NOC4 NOP14 NSA2 NTH1 NUP OGG1 OLE2 OPI1 OPT1 OPT6 OPT9 ORC1 ORC3 ORC4 OSM2 OYE23 PAD1 PCK1 PCL5 PDC12 PDX1 PEP7 PEX11 PEX13 PEX14 PGA13 PGA14 PGA23 PGA25 PGA39 PGA41 PGA45 PGA60 PHA2 PHM7 PHO23 PHR1 PIL1 PLB2 PLC1 PLC2 PMM1 PMT6 POL93 POT1 POX1-3 PPH21 PR26 PRE5 PRE8 PRN2 PRN3 PRN4 PRP22 PRP39 PRR2 PRT1 PSP1 PST1 PSY2 PTC5 PTC7 PTH1 PTH2 PUF3 PUP3 PUS4 PUT1 PUT2 PUT4 PWP2 PXA1 PXA2 PXP2 PYC2 RAD54 RBR1 RCE1 RCF3 RET2 RFC1 RFC2 RFC4 RFX2 RGS2 RHD2 RIX7 RLI1 RLM1 RMS1 RPB4 RPL7 RPN10 RPN5 RPN8 RPO26 RPS27 RPT6 RRN3 RRP15 RRS1 RSR1 RTA4 SAP2 SAP8 SAP99 SAS3 SCL1 SDC1 SDH4 SEC14 SEC4 SEC65 SEF1 SEF2 SEN1 SEO1 SER33 SET3 SFL2 SFU1 SGA1 SGT1 SIR2 SIT1 SKI8 SLN1 SMC2 SMD3 SMF12 SNM1 SNU114 SOD1 SOD2 SOD3 SOD4 SOL3 SOU1 SPC3 SPC98 SPE2 SPO75 SPR3 SSC1 SSP96 SST2 SUI3 SUL2 SUP35 SUR7 SVF1 SWD2 SWI1 TAF145 TEA1 TES1 TES15 THI13 THR1 THR4 THS1 TIF5 TIP1 TIP20 TLG2 TLO16 TPK2 TPO4 TRM2 TRP2 TRP4 TRP99 TRR1 TRX2 TSR1 TUB4 TYS1 UAP1 UBC15 UGA33 UGA6 ULP1 ULP3 URA2 URK1 USO5 UTP13 UTP21 UTP22 UTP5 UTP9 VAC8 VCX1 VID27 VMA22 VPH1 VPS1 VPS16 VPS2 VPS4 WH11 YBL053 YBN5 YHB5 YNK1 YTM1 YUH2 ZCF15 ZCF22 ZCF25 ZCF27 ZCF4 ZCF5 ZCF9 ZNC1 ZRT1 ZWF1

Annexe G1: Stepwise filtering of read pairs in the control strain (T0) and the four derived mistranslating strains before genotype calling.

		T0	T1	T2	T2KO1	T2KO2
1	**Raw reads**	6,141,944	8,437,847	10,336,605	6,923,919	11,730,542
2	**Reads ready to align** [1)]	5,125,425	6,975,400	8,552,088	5,708,627	6,012,173
3	**Filtering duplications**	3,256,140	4,362,498	5,376,171	3,561,454	5,933,352
4	**Filtering improper reads**	3,190,472	4,267,283	5,266,332	3,487,909	5,773,977
5	**Filtering unmapped reads**	3,084,224	4,115,782	5,111,017	3,362,403	5,546,510
6	**Mapped reads ready to analysis for genotype calling** [2)]	2,978,819	3,976,344	4,934,771	3,266,126	5,412,111

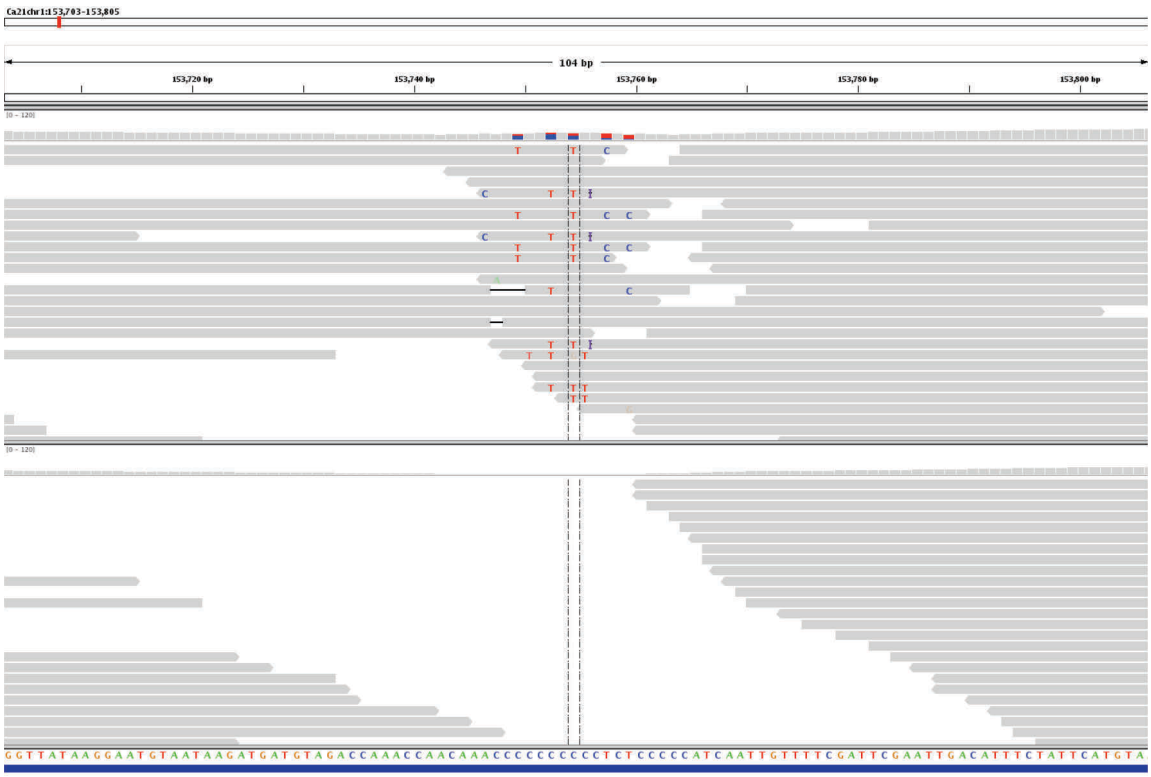
1) after filtering reads for ambiguity

2) after filtering dodgy regions

Annexe G2: Average alignment depth and reference genome coverage of mapping. Total coverage of mapped reads before and after filtering regions not well covered by the reference genome, calculated as the average depth of reads mapped to the reference genome.

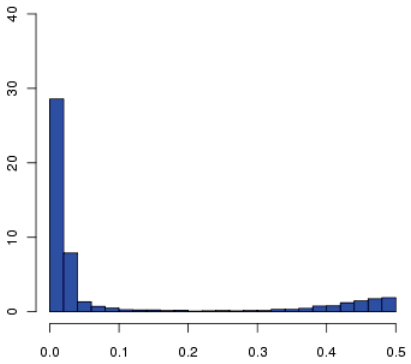
	T0	T1	T2	T2KO1	T2KO2
Average alignment depth before removel dodgy regions	59.9	79.2	99.8	64.6	93.5
Average alignment depth after removel dodgy regions	56.8	76.4	96.2	62.7	91.0
Reference genome coverage before removel dodgy regions (%)	99.8	99.8	99.8	99.7	99.8
Reference genome coverage after removel dodgy regions (%)	99.4	99.4	99.4	99.4	99.6

Annexe G3: Alignment of reads from the T0 control sample ambiguously mapping to a genomic region on chromosome 1 in the reference genome SC5314. One large fraction of the reads had perfect sequence identity with the reference regions, whereas another portion had frequent and identical differences to this region, indicating a different sequence source not present in the reference genome. These ambiguous regions were filtered out. The lower half of the figure shows the same region after filtering.

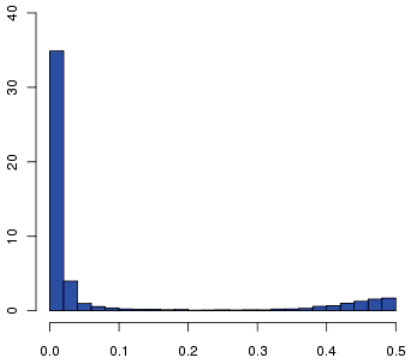


Annexe G4: The distribution of minor allele frequency (MAF) for bi-allelic positions in the five studied strains. The estimated MAF at polymorphic locations in each sample shows a distribution dominated by low frequencies (MAF <= 10%), caused by sequencing errors and misalignment, and high frequencies (MAF >= 40%), caused by heterozygotes.

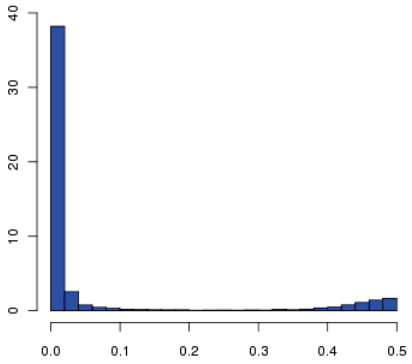
T0



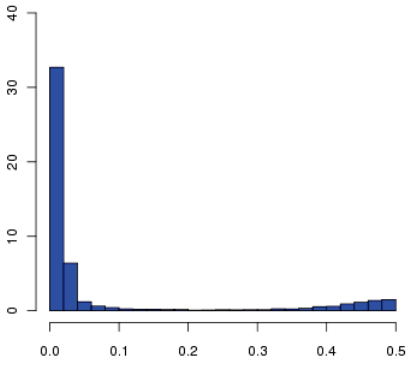
T1



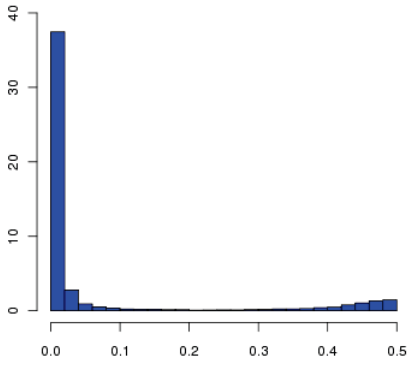
T2



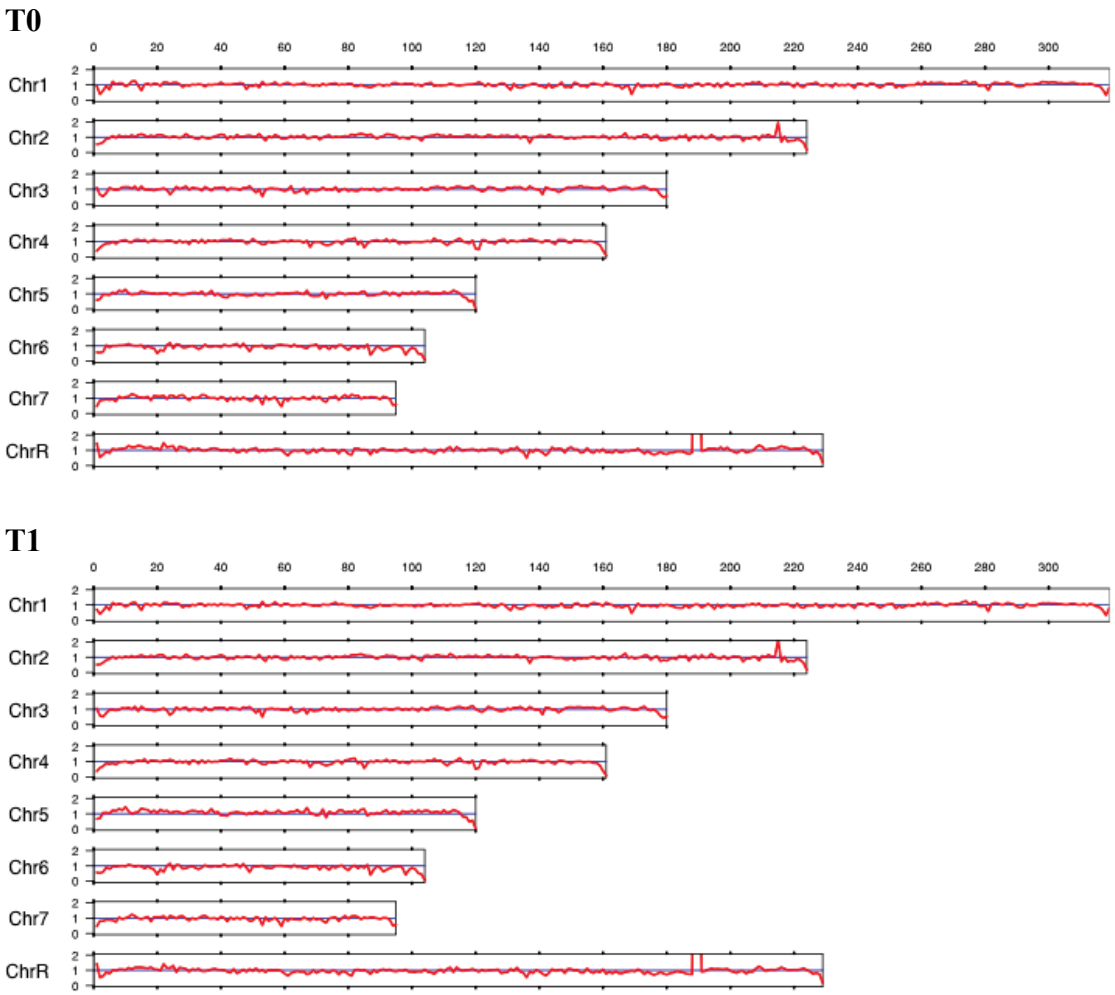
T2KO1



T2KO2

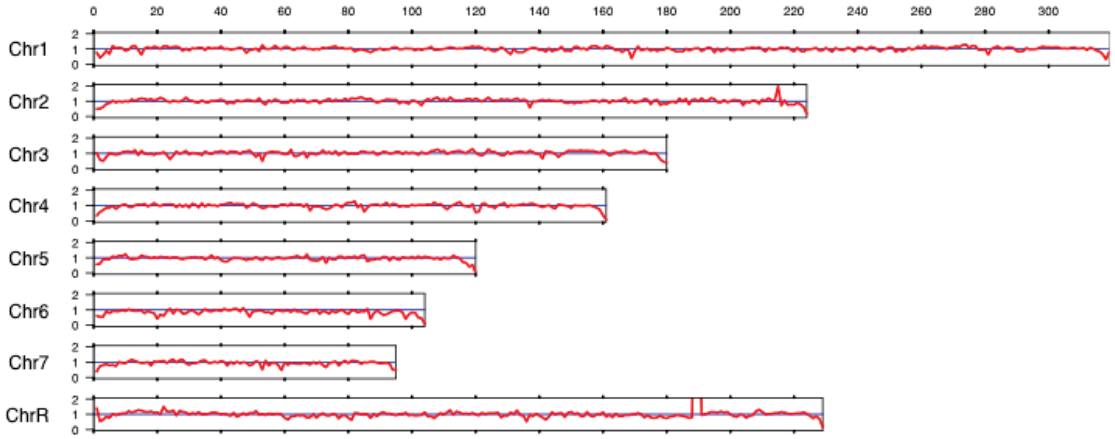


Annexe G5: Read coverage is relatively uniform across the genome. For each sample (T0, T1, T2, T2KO1 and T2KO2), average read depth in a sliding window of 10 kilobases along each chromosome was normalized to the average depth for the whole corresponding sample. Apart from a short section on ChrR in all samples, the coverage is relatively uniform for all samples. The decreasing number towards the chromosome ends is caused by numeric effects from the sliding window.

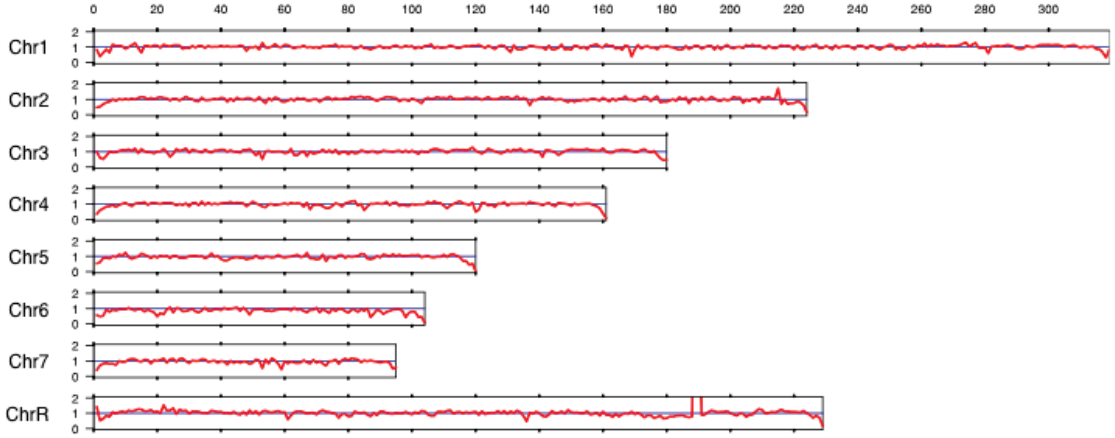


Annexe G5: Read coverage is relatively uniform across the genome.

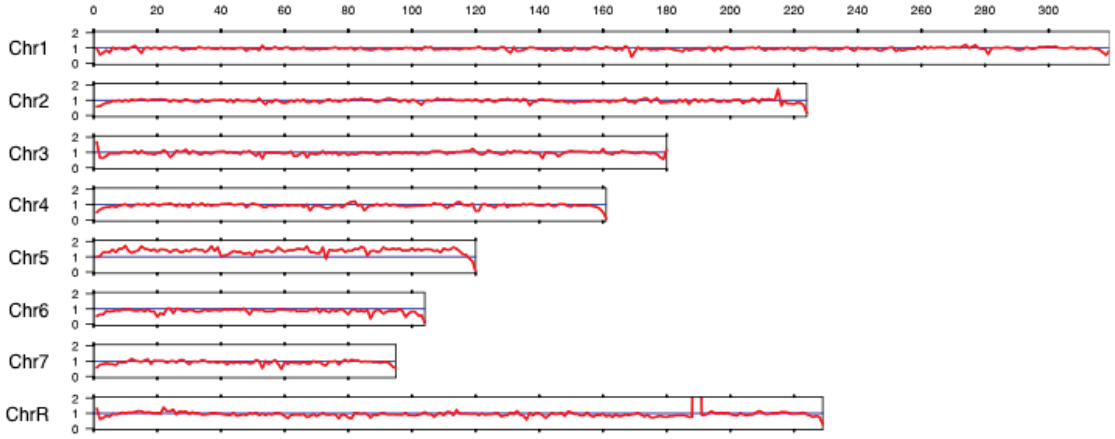
T2



T2KO1

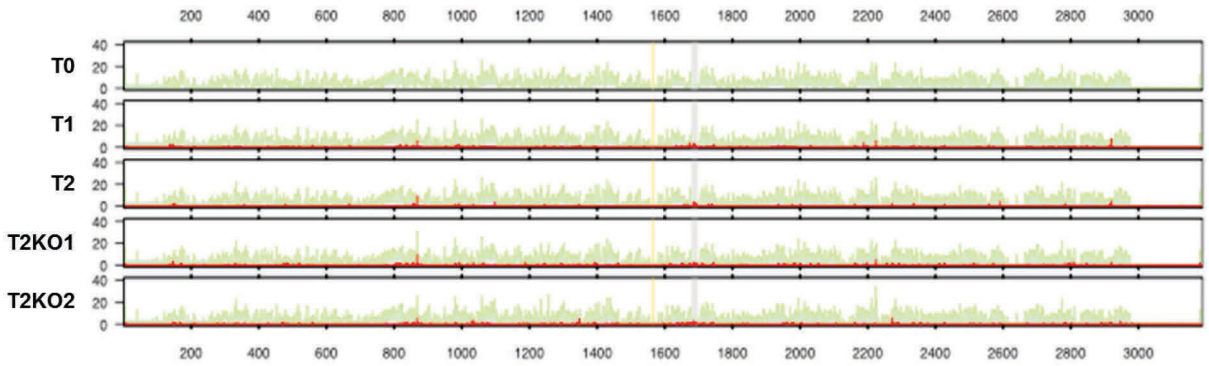


T2KO2

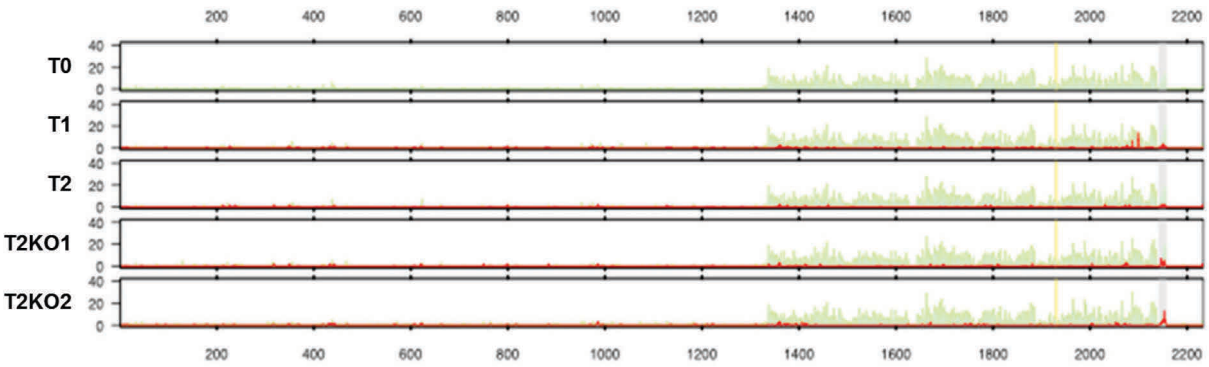


Annexe G6: The SNP density in the five sequenced strains. The SNP density (SNPs/kilobase, shaded green) varies along the chromosomes (shown as subfigures) but are essentially similar between the five strains (represented by tracks within each subfigure). There are homozygous stretches matching in all samples, as well. The density of SNPs with loss of heterozygosity (LOH) in a derived strain compared to T0 are plotted in red, showing a region on Chromosome 5 in both T2KO1 and T2KO2, and the entire Chromosome R in T2KO1, where all heterozygote SNPs have changed to homozygous. Yellow and grey vertical lines indicate centromere and MRS, respectively.

Chr1

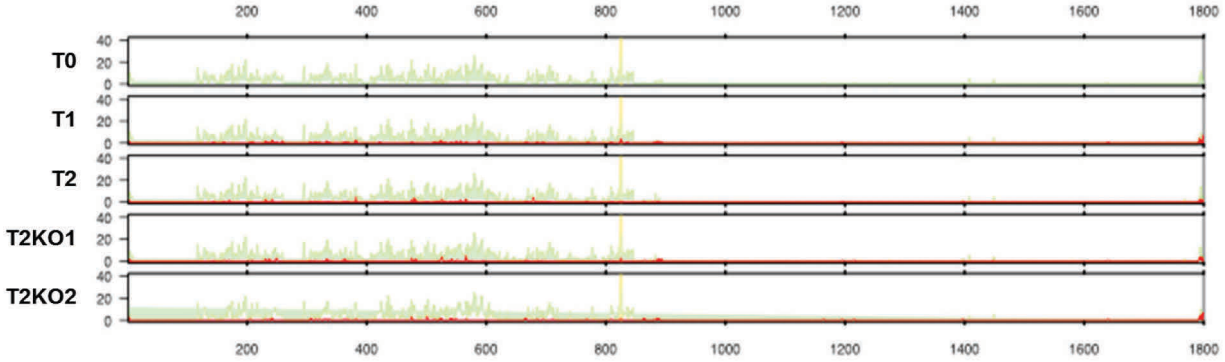


Chr2

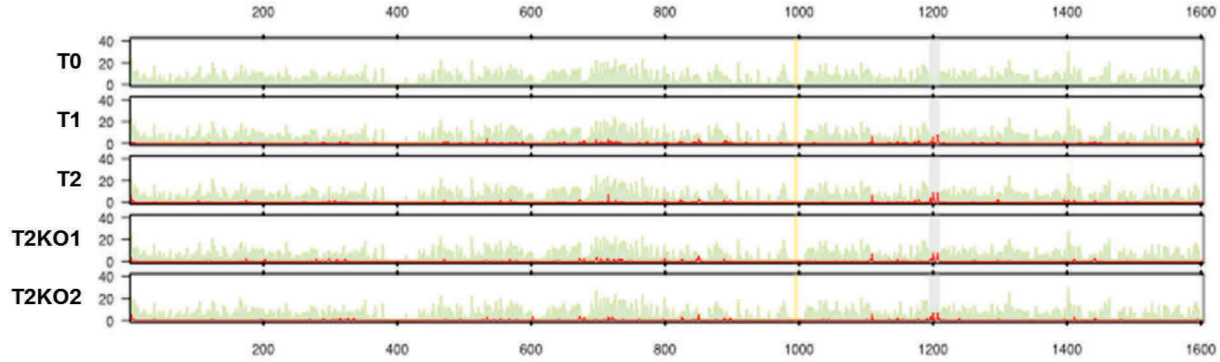


Annexe G6: The SNP density in the five sequenced strains.

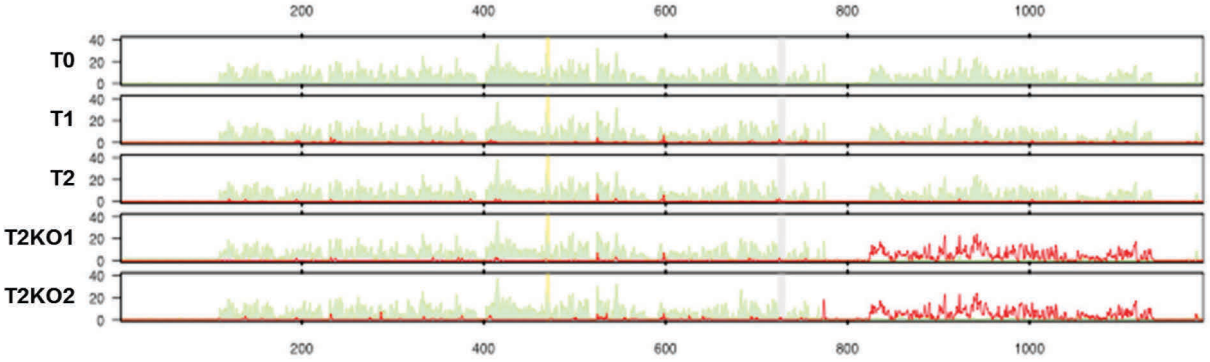
Chr3



Chr4

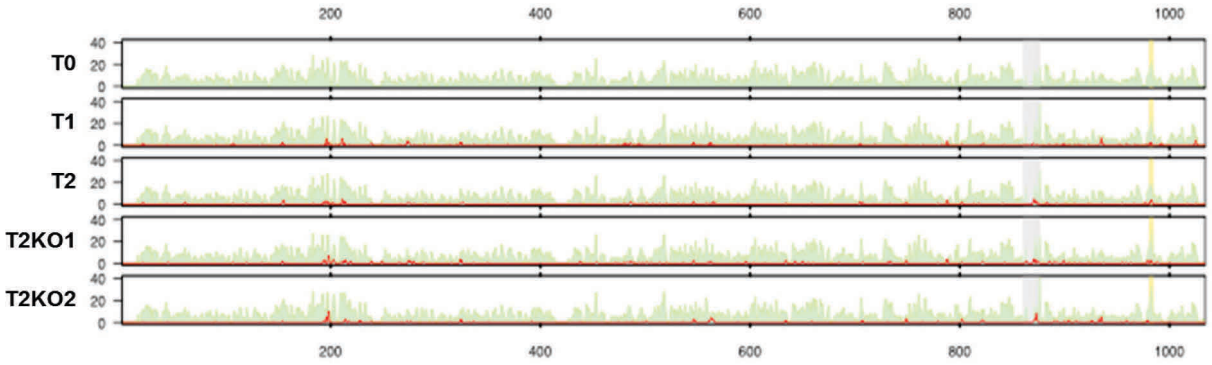


Chr5

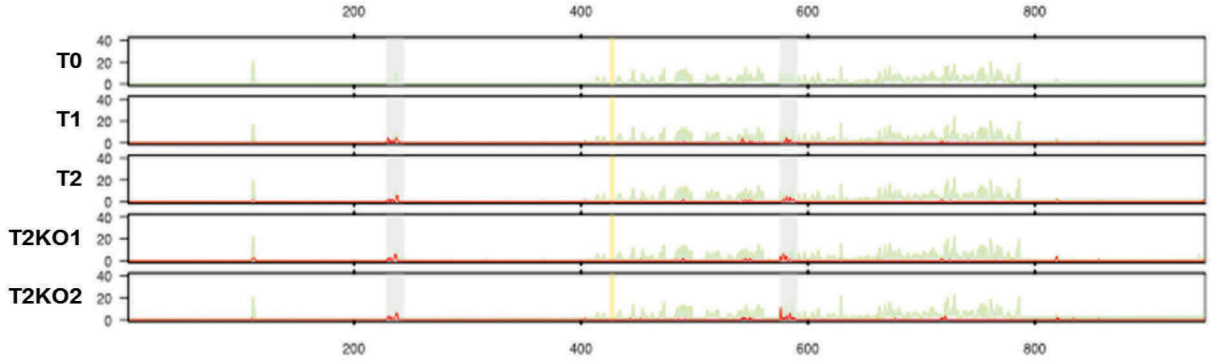


Annexe G6: The SNP density in the five sequenced strains.

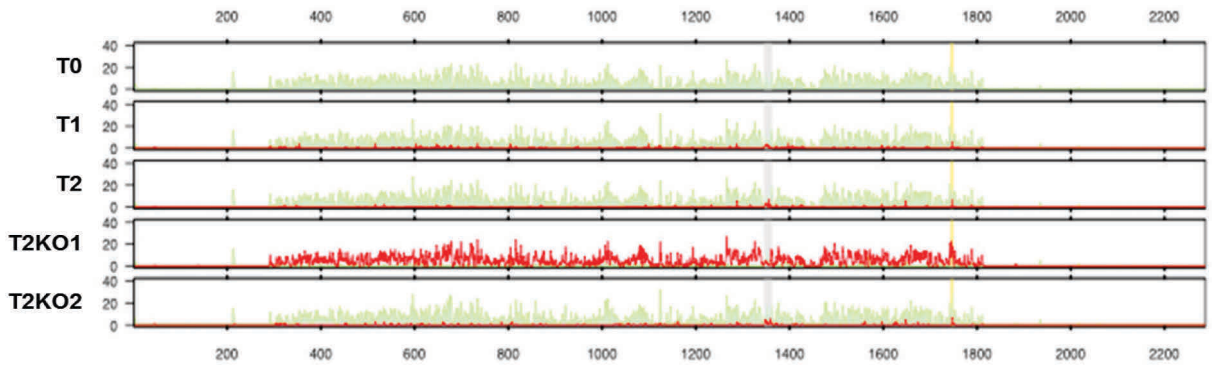
Chr6



Chr7



ChrR



Annexe G7: Loss of heterozygosity (LOH) from T0

	T1_LOH	T2_LOH	T2KO1_LOH	T2KO2_LOH
A. LOH of each sample from T0				
chr1	226	218	266	189
chr2	159	130	153	156
chr3	86	84	92	77
chr4	169	173	185	146
chr5	90	90	1768	1809
chr6	154	141	168	140
chr7	56	61	70	75
chrR	188	146	8137	173
all	1128	1043	10839	2765
B. LOH normalized by total number of SNPs in T0				
chr1	0.41	0.4	0.49	0.35
chr2	0.29	0.24	0.28	0.29
chr3	0.16	0.15	0.17	0.14
chr4	0.31	0.32	0.34	0.27
chr5	0.16	0.16	3.23	3.31
chr6	0.28	0.26	0.31	0.26
chr7	0.1	0.11	0.13	0.14
chrR	0.34	0.27	14.87	0.32
all	2.06	1.91	19.81	5.05
C. LOH normalized by SNPs number of each chromosome in T0				
chr1	1.46	1.41	1.72	1.23
chr2	3.41	2.79	3.28	3.34
chr3	2.34	2.28	2.50	2.09
chr4	2.07	2.12	2.27	1.79
chr5	1.46	1.46	28.78	29.44
chr6	2.32	2.12	2.53	2.10
chr7	3.54	3.85	4.42	4.74
chrR	2.23	1.73	96.51	2.05

Annexe G8: Validation of LOH in both T2KO1 and T2KO2 with different homozygotes using Sanger sequencing.

	Chr_Pos	T0	T2KO1	T2KO2	Sanger	ORF_ID	GeneSymbol
HT_1	Ca21chr5_110754	C/A	C/A	C/A	confirmed	orf19.935	AGA1
HT_2	Ca21chr5_175512	A/G	A/G	A/G	confirmed		
HT_3	Ca21chr5_244781	G/A	G/A	G/A	confirmed	orf19.1960	CLN3
HT_4	Ca21chr5_254717	A/G	A/G	A/G	confirmed		
HT_5	Ca21chr5_277985	T/C	T/C	T/C	confirmed	orf19.1945	AUR1
HT_6	Ca21chr5_338426	C/T	C/T	C/T	confirmed	orf19.4146	SMD3
HT_7	Ca21chr5_428763	C/T	C/T	C/T	confirmed	orf19.3183	
HT_8	Ca21chr5_535749	A/T	A/T	A/T	confirmed	orf19.4251	ZCF22
LOH_8	Ca21chr5_898285	G/A	G/G	A/A	confirmed		
LOH_9	Ca21chr5_905792	G/A	G/G	A/A	confirmed	orf19.1283	MEC1
LOH_10	Ca21chr5_917859	C/T	T/T	C/C	confirmed	orf19.3893	SCW11
LOH_11	Ca21chr5_947363	A/G	A/A	G/G	confirmed		
LOH_12	Ca21chr5_977111	A/G	A/A	G/G	confirmed	orf19.3931	SFC1
LOH_13	Ca21chr5_1074638	G/A	A/A	G/G	confirmed	orf19.3981	MAL31
LOH_14	Ca21chr5_1114909	A/C	A/A	C/C	confirmed		
LOH_15	Ca21chr5_1133037	G/A	A/A	G/G	confirmed	orf19.4009	CNB1

Annexe G9: Genotype (GT) changes. Genotype changes in the four mistranslating strains compared to the control strain T0.

	T0 - T1	T0 - T2	T0 - T2KO1	T0 - T2KO2
No. of GT changes	2,676	2,627	12,582	5,211
No. of GT changes excluding two regions [1)]	2,212	2,177	2,286	2,879
No. of GT changes (non-coding) [1)]	1,507	1,459	1,496	1,924
No. of GT changes (coding) [1)]	705 (31.9%)	718 (33%)	790 (34.6%)	955 (33.2%)
No. of GT changes (non-synonymous) [1)]	312 (14.1%)	297 (13.6%)	314 (13.7%)	399 (13.9%)
No. of GT changes (synonymous) [1)]	311	330	382	443

1) Chr5: 825000 ~ 1136000
ChrR: 291000 ~ 1814000

Annexe G10: Genes with nonsynonymous GT changes only in T2KO2 sample.

Chr.	GeneID	Symbol	Chr.	GeneID	Symbol	Chr.	GeneID	Symbol
Ca21chr1	orf19.1043		Ca21chr2	orf19.192		Ca21chr4	orf19.5041	
Ca21chr1	orf19.1228	HAP2	Ca21chr2	orf19.3148		Ca21chr4	orf19.5290	
Ca21chr1	orf19.3336		Ca21chr2	orf19.3490	FGR6-4	Ca21chr4	orf19.738	MYO5
Ca21chr1	orf19.3685	PSY2	Ca21chr2	orf19.3550	SGO1	Ca21chr5	orf19.1935	
Ca21chr1	orf19.3686		Ca21chr2	orf19.3599	TIF4631	Ca21chr5	orf19.1942	SGE1
Ca21chr1	orf19.4061		Ca21chr2	orf19.3600		Ca21chr5	orf19.2650	
Ca21chr1	orf19.4472		Ca21chr2	orf19.5820	UGA6	Ca21chr5	orf19.3205	
Ca21chr1	orf19.4753	PFK26	Ca21chr2	orf19.6866		Ca21chr5	orf19.4054	CTA24
Ca21chr1	orf19.4757	NAR1	Ca21chr3	orf19.1288	FOX2	Ca21chr5	orf19.4216	
Ca21chr1	orf19.48		Ca21chr3	orf19.1606		Ca21chr5	orf19.4280	
Ca21chr1	orf19.4921		Ca21chr3	orf19.1611		Ca21chr5	orf19.4301	
Ca21chr1	orf19.4961	STP2	Ca21chr3	orf19.1616	FGR23	Ca21chr5	orf19.4312	
Ca21chr1	orf19.4964		Ca21chr3	orf19.1656		Ca21chr5	orf19.4318	MIG1
Ca21chr1	orf19.4967	COX19	Ca21chr3	orf19.229		Ca21chr5	orf19.4330	
Ca21chr1	orf19.5068	IRE1	Ca21chr3	orf19.245	DDC1	Ca21chr6	orf19.100	
Ca21chr1	orf19.6020		Ca21chr3	orf19.2515		Ca21chr6	orf19.1077	ATM1
Ca21chr1	orf19.6022		Ca21chr3	orf19.313	DAL4	Ca21chr6	orf19.115	
Ca21chr1	orf19.6027		Ca21chr3	orf19.5869		Ca21chr6	orf19.1214	
Ca21chr1	orf19.6242	CYK3	Ca21chr4	orf19.1234	FGR6-10	Ca21chr6	orf19.3429	FGR47
Ca21chr1	orf19.6244		Ca21chr4	orf19.1798	TSC2	Ca21chr6	orf19.3448	
Ca21chr1	orf19.6261	BPH1	Ca21chr4	orf19.2743		Ca21chr6	orf19.3475	
Ca21chr1	orf19.6277		Ca21chr4	orf19.2901	NUP60	Ca21chr6	orf19.4555	ALS4
Ca21chr2	orf19.1348		Ca21chr4	orf19.3362		Ca21chr6	orf19.5537	WSC2
Ca21chr2	orf19.1491		Ca21chr4	orf19.3370	DOT4	Ca21chr6	orf19.5543	
Ca21chr2	orf19.151	TPO5	Ca21chr4	orf19.3794	CSR1	Ca21chr6	orf19.5718	
Ca21chr2	orf19.1539		Ca21chr4	orf19.4643		Ca21chr6	orf19.5742	ALS9
Ca21chr2	orf19.1893		Ca21chr4	orf19.4690		Ca21chr6	orf19.5775	
Ca21chr6	orf19.99	HAL21	Ca21chr7	orf19.5191	FGR6-1	Ca21chr7	orf19.6453	
Ca21chr7	orf19.6456		Ca21chr7	orf19.6486	LDG3			