The genome of Candida cylindracea to study the role of the translational machinery on gene evolution

Rita Coimbra[1,2], Miguel Pinheiro[1], Andreia Reis[1], Ana R. Bezerra[1], José Luís Oliveira [2], Jean-Luc-Souciet [3], Manuel A.S. Santos[1] and Gabriela Moura[1]

[1] Department of Medical Sciences & iBiMED, University of Aveiro; [2] Department of Electronics and Telematics, IEETA, University of Aveiro; [3] Institute de Botanique, Strasbourg, France.

The genetic code is degenerate allowing for introduction of species-specific codon utilization biases. These has been intensively studied but it is not yet clear whether they arise from mutational pressure alone or whether other forces, such as the translational efficiency, also influence it. To address this question we took advantage of a unique genetic code alteration that occurred in the fungal Saccharomycotina subphylum, the so-called CTG clade, which translate leucine CTG codons as serine due to a novel transfer RNA (tRNACAGSer). The atypical serine CTG codons are rarely used among the CTG clade species except for *Candida cylindracea* which uses CTGs at high level, allowing to gain further insight on how they were remodeled after the reassignment event.

For this, we conducted *de novo* whole genome sequencing, and annotation of *C. cylindracea*. Alongside, we performed systematic alignment of yeast orthologous genes, phylogenetic reconstruction and codon evolution analysis. The results showed that *C. cylindracea*'s genome is different from the other Saccharomycotina genomes and occupies a basal position with respect to the reassignment event. Our findings support the hypothesis that CTG codon reassignment, and not only G+C pressure, exerted a strong influence on serine/leucine codon evolution in yeast.

Acknowledgments

This work is supported by the Portuguese Foundation for Science and Technology (FCT) and European Regional Development Fund (FEDER), through COMPETE2020 (UID/BIM/04501/2013; Centro-45-2015-01; PTDC/BIA-MIC/31849/2017,). Rita Coimbra is supported by the Project 016385 - PAC through the grant BPD/UI62/7535/2017.

sciences

The genome of Candida cylindracea to study the role of the translational machinery on gene evolution

Rita Coimbra[1,2],
Miguel Pinheiro[1], Andreia Reis[1], Ana R. Bezerra[1], José Luís Oliveira [2], Jean-Luc-Souciet [3], Manuel A.S. Santos[1] and Gabriela Moura[1]

[1] Department of Medical Sciences & iBiMED, University of Aveiro; [2] Department of Electronics and Telematics, IEETA, University of Aveiro; [3] Institute de Botanique, Strasbourg, France.

Abstract

The genetic code is degenerate allowing for introduction of species-specific codon utilization biases. These has been intensively studied but it is not yet clear whether they arise from mutational pressure alone or whether other forces, such as the translational efficiency, also influence it. To address this question we took advantage of a unique genetic code alteration that occurred in the fungal Saccharomycotina subphylum, the so-called CTG clade, which translate leucine CTG codons as serine due to a novel transfer RNA (tRNACAGSer). The atypical serine CTG codons are rarely used among the CTG clade species except for *Candida cylindracea* which uses CTGs at high level, allowing to gain further insight on how they were remodeled after the reassignment event.
For this, we conducted *de novo* whole genome sequencing, and annotation of *C. cylindracea*. Alongside, we performed systematic alignment of yeast orthologous genes, phylogenetic reconstruction and codon evolution analysis. The results showed that *C. cylindracea*'s genome is different from the other Saccharomycotina genomes and occupies a basal position with respect to the reassignment event. Our findings support the hypothesis that CTG codon reassignment, and not only G+C pressure, exerted a strong influence on serine/leucine codon evolution in yeast.

Introduction

Now we know that the yeast CTG reassignment event has a polyphyletic nature [1] and erased thousands of codons from the genome of the CTG clade ancestors [2]. As a consequence, these atypical serine CTG codons are rarely used among the CTG clade species, and the $tRNA_{CAG}$Ser that decodes it is also a low abundance tRNA, normally encoded by one single gene copy. *Candida cylindracea* is the exception to the rule as it uses CTGs at high level (Fig.1), and has at least 3 genes for this tRNA (data not shown). We have sequenced the entire genome of *C. cylindracea* to gain further insight on how and why the usage of this codon diverged so dramatically between *C. cylindracea* and the other CTG clade species.

Methods

Whole genome sequencing and assembling was conducted at Genolevures Program, scaffolds annotation was done using MAKER [3]. To create the phylogenetic tree, we used 28 proteomes from PhylomedB together with *Candida cylindracea*. We ran OrthoMCL algorithm to identify the orthologous, chosing the ones with more than 50% match between pairs of proteins. Within orthologous sets we isolated groups that had hits in all species using *Saccharomyces cerevisiae* as root. With this method we obtained 357 groups of orthologous that were present in 29 species. We aligned these 357 groups separately with T-Coffee (11.0.8) using Mafft, Muscle and Kalign methods. We ran the Trimal (1.2rev59) in all alignments to trim regions with less than 80% match. The trimmed alignments were concatenated yielding a total sequence of 184718bp. Smart Model Selection in PhyML (SMS 1.8.1) was then used to create the phylogenetic tree. SMS selects the substitution model LG'+G+I+F' with AIC=12423869.84956. The codons present in each aligned position for all species, at CTG positions, were computed through in-house made Python scripts.

Conclusions

C. cylindracea prefers GC-rich codons and uses CTG as the preferred serine codon (Fig.1).

According to the phylogeny reconstruction, *C. cylindracea* has a basal position in relation to the other non-standard yeast decoders (Fig.2). The phylogenetic tree places *C. cylindracea* at the margin of the CTG-Ser decoders, closer to the Ser2 clade [4].

Although highly abundant, practically none of the CTG codons of *C. cylindracea* are conserved at the codon level, due to the reassignment of this codon from leucines to serines (relative to the standard decoders) and to the ambiguity of its decoding (still present in many CTG clade species).

Ongoing lines of research aim at confirming the evolutionary origin of *C. cylindracea*'s $tRNA_{CAG}$Ser genes and if CTG usage influences gene expression, which could explain its intriguing generalized use in *C. cylindracea*'s genome.

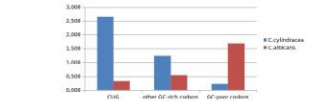

Fig 1. ***C. cylindracea* CTG usage.** The relative synonymous codon usage (RSCU) of serine codons has been calculated using *C. cylindracea* and *C.albicans* ORFeomes. CTG usage has been plotted separately while the other serine codons were divided into GC-rich codons (average RSCU of AGC, UCG and UCC is shown) and GC-poor codons (average RSCU of AGU, UCA and UCU is shown). *C. cylindracea* prefers GC-rich codons but the usage of the CTG codon is twice as high has that of other comparable GC-rich serine codons, suggesting that the CTG usage has been influenced by other evolutionary pressures beyond GC content.

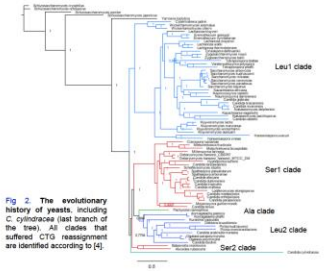

Fig 2. **The evolutionary history of yeasts**, including *C. cylindracea* (last branch of the tree). All clades that suffered CTG reassignment are identified according to [4].

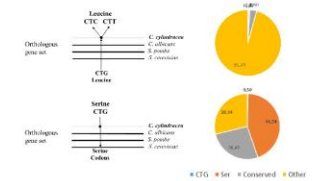

Fig 3. **CTG remodeling in *C. cylindracea*.** The alignments of orthologous proteins were analyzed focusing on the degree of conservation of CTG codons along the species. Interestingly, only less than 1% of *C. cylindracea* extant CTG codons were conserved as CTG codons in the other species (below), suggesting a major remodeling of CTG usage among yeasts. Nevertheless, CTGs behave as serine codons in *C. cylindracea*, since most of this positions are conserved for serine residues (44%). When CTG positions were fixed in the other species (above), they were massively mutated into other type of amino acid, on the *C. cylindracea* side, mostly into Leu-CTC and Leu-CTT.

References

[1] Mühlhausen, S. and Kollmar, M. 2014. Genome Biol. Evol. 6: 3222-3237.
[2] Massey, S. et al. 2003. Genome Research. 13. 544-557.
[3] http://www.yandell-lab.org/software/maker.html
[4] Krassowski et al. 2018. Nature Communications. 9: 1887.

Acknowledgments

This work is supported by the Portuguese Foundation for Science and Technology (FCT) and European Regional Development Fund (FEDER), through COMPETE2020 (UID/BIM/04501/2013; Centro-45-2015-01; PTDC/BIA-MIC/31849/2017,). Rita Coimbra is supported by the Project 016385 - PAC through the grant BPD/UI62/7535/2017.