

Clark University

Clark Digital Commons

Undergraduate Student Research Festivals

Academic Spree Day 2021

Apr 26th, 12:00 AM

Comparative Analysis of Deviations from the Genetic Code in Yeasts

Natalie Zaba

Clark University, nzaba@clarku.edu

Follow this and additional works at: <https://commons.clarku.edu/asdff>

Zaba, Natalie, "Comparative Analysis of Deviations from the Genetic Code in Yeasts" (2021).

Undergraduate Student Research Festivals. 23.

<https://commons.clarku.edu/asdff/asd2021/asd2021/23>

This Open Access Event is brought to you for free and open access by the Conference Proceedings at Clark Digital Commons. It has been accepted for inclusion in Undergraduate Student Research Festivals by an authorized administrator of Clark Digital Commons. For more information, please contact larobinson@clarku.edu, mkrikonis@clarku.edu.

Comparative Analysis of Deviations from the Genetic Code in Yeasts

Natalie Zaba, David S. Hibbett, and Miguel A. Naranjo-Ortiz



CLARK UNIVERSITY

Introduction

- Every living organism follows the genetic code (Image 1), which is a code that tells the cell how to make proteins based on RNA sequences.

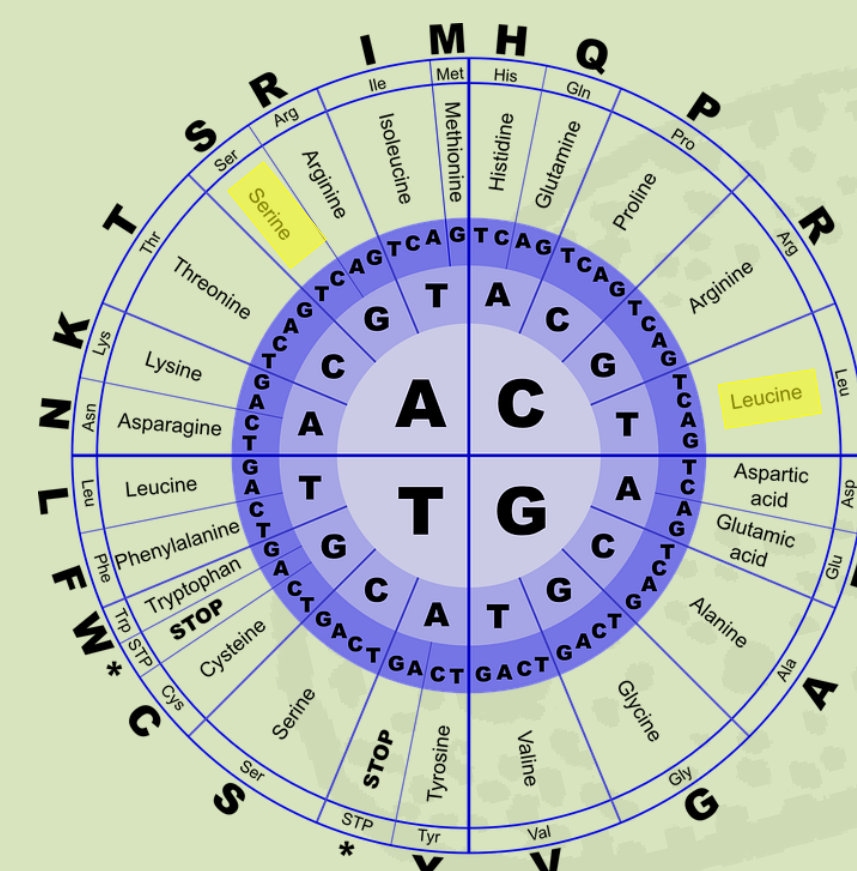


Image 1: The Genetic Code

Swift, David W. "The Genetic Code." *Evolution under the Microscope - Origin of Life*, Apr. 2017, evolutionunderthemicroscope.com/ool02.html.

- This code is nearly universal, but many deviations are known. In yeast, codon reassignment has evolved at least three times independently. The CTG codon normally encodes for Leucine, but in two yeast groups it's translated into serine and one into alanine.
- Candida albicans* is the most studied species that has this reassignment.

- Candida* species are the most common human fungal pathogens and are responsible for both superficial and life-threatening systemic infections.
- Certain species are used in the production of bread, beer and wine
- There are approximately 1500 described species of Saccharomycotina, the main group of yeast fungi, which have great metabolic and genetic diversity.

Goals

- See if genes with a higher CTG score are linked to more important cellular functions such as reproduction, growth and potentially virulence.
- Understanding functional implication of CTG codon reassignment in yeasts.

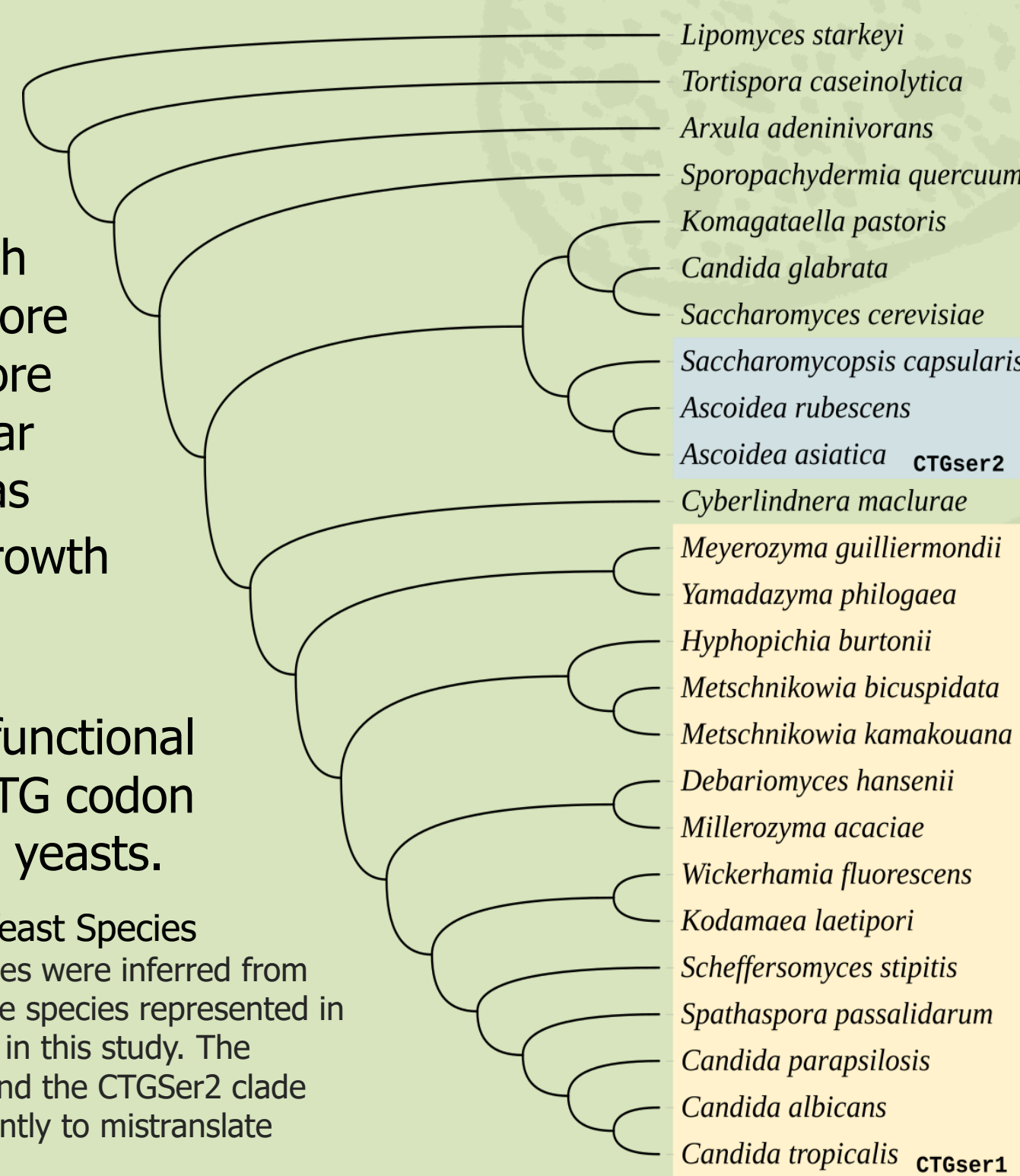


Fig. 1: Phylogeny of Yeast Species

The topology of the species were inferred from Shen et al. (2018). All the species represented in the phylogeny were used in this study. The CTGSer1 clade (yellow) and the CTGSer2 clade (blue) evolved independently to mistranslate Leucine for Serine

Table 1: Most Common Biological Processes in all Species

Term	Process	Count
GO:0036211	Protein modification process	21
GO:0006464	Cellular protein modification process	21
GO:0065007	Biological regulation	21
GO:0090304	Nucleic acid metabolic process	20
GO:0016070	RNA metabolic process	20
GO:0043412	Macromolecule modification	20
GO:0006468	Protein phosphorylation	19
GO:0043170	Macromolecule metabolic process	19
GO:0044260	Cellular macromolecule metabolic process	18
GO:0010556	Regulation of macromolecule biosynthetic process	18
GO:0031326	Regulation of cellular biosynthetic process	18
GO:2000112	Regulation of cellular macromolecule biosynthetic process	18
GO:0009889	Regulation of biosynthetic process	18

Table 3: Most Common Biological Processes in CTGSer1 and CTGSer2

GO Term	Biological Process	CTGSer1 Count	CTGSer2 Count
GO:0006464	Cellular protein modification process	11	3
GO:0036211	Protein modification process	11	3
GO:0006468	Protein phosphorylation	10	3
GO:0043412	Macromolecule modification	10	2
GO:0015985	Energy coupled proton transport, down electrochemical gradient	9	1
GO:0065007	Biological regulation	9	3
GO:0072522	Purine-containing compound biosynthetic process	9	1
GO:0090304	Nucleic acid metabolic	9	2
GO:0016070	RNA metabolic process	9	2
GO:0009199	Ribonucleoside triphosphate metabolic process	9	1
GO:0009142	Nucleoside triphosphate biosynthetic process	9	1
GO:0009205	Purine ribonucleoside triphosphate metabolic process	9	1
GO:0009144	Purine nucleoside triphosphate metabolic process	9	1
GO:0015986	ATP synthesis coupled proton transport	9	1
GO:0006754	ATP biosynthetic process	9	1
GO:0009201	Ribonucleoside triphosphate biosynthetic process	9	1
GO:0009145	Purine nucleoside triphosphate biosynthetic process	9	1
GO:0009206	Purine ribonucleoside triphosphate biosynthetic process	9	1

The sample of genes used from each species were based on the 10th percentile of the CTG score. The CTG score was created in order to determine the "CTG-ness" of a gene. The CTG score is: $-\log(1 + \text{total number of "CTG"}) / \text{length of sequence}$. We assigned Gene Ontology terms to the genes of all species using Wei2Go. GO describes the biology of a gene in three different aspects: Biological Process, Cellular Component, and Molecular Function. A sample of 25 species were used were used. The CTGSer1 clade had 14 species in the sample, and the CTGSer2 clade which had 3 species in the sample. Table 1 shows the most common biological processes among all the species. The number of species that have each biological process appears is displayed under "Count" in the table. Table 2 shows the most common molecular function among all the species. Like Table 1, Table 2 shows the count of how many times each molecular process appeared. Table 3 shows the most common biological processes in the CTGSer1 and CTGSer2 clades. The appearance of the processes and functions were recorded under "Count." Table 2 shows the most common molecular functions in both clades.

Table 2: Most Common Molecular Functions in all Species

Term	Function	Count
GO:0032559	Adenyl ribonucleotide binding	21
GO:0005524	ATP binding	21
GO:0003676	Nucleic acid binding	21
GO:0030554	Adenyl nucleotide binding	20
GO:0035639	Purine ribonucleoside triphosphate binding	20
GO:0003677	DNA binding	20
GO:0016787	Hydrolase activity	20
GO:0005488	binding	20
GO:0004672	Protein kinase activity	19
GO:0016301	Kinase activity	19
GO:0032555	Purine ribonucleotide binding	19
GO:1901363	Heterocyclic compound binding	19
GO:0043167	Ion binding	19
GO:0097159	Organic cyclic compound binding	19

Table 4: Most Common Molecular Functions in CTGSer1 and CTGSer2

GO Term	Molecular Function	CTGSer1 Count	CTGSer2 Count
GO:0003677	DNA binding	10	0
GO:0005524	ATP binding	10	3
GO:0030554	Adenyl nucleotide binding	10	3
GO:0032559	Adenyl ribonucleotide binding	10	3
GO:0004672	Protein kinase activity	10	3
GO:0016301	Kinase activity	10	3
GO:0003674	Molecular function	9	0
GO:0005488	Binding	9	3
GO:0016740	Transferase activity	9	0
GO:0005515	Protein binding	9	3
GO:0043167	Ion binding	9	0
GO:0097367	Carbohydrate derivative binding	9	0
GO:1901363	Heterocyclic compound binding	9	0
GO:0097159	Organic cyclic compound binding	9	0
GO:0003676	Nucleic acid binding	9	3
GO:0032553	Ribonucleotide binding	9	0
GO:0032555	Purine ribonucleotide binding	9	3
GO:0016772	Transferase activity, transferring phosphorus-containing groups	9	0
GO:0043168	Anion binding	9	0
GO:0035639	Purine ribonucleoside triphosphate binding	9	3
GO:0017076	Purine nucleotide binding	9	0
GO:0016491	Oxidoreductase activity	6	3
GO:0003723	RNA binding	3	3
GO:0140110	Transcription regulator activity	5	3
GO:0005975	Carbohydrate metabolic process	0	3
GO:0006468	Protein phosphorylation	0	3
GO:0016787	Hydrolase activity	9	0

Results & Discussion

- The most common biological processes among all the species are protein modification, cellular protein modification and biological regulation (Table 1).
- The most common molecular functions among all the species are adenyl ribonucleotide binding, ATP binding, and adenyl nucleotide binding (Table 2).
- Energetic metabolism and ATP biosynthesis seems to be enriched in the CTGSer1 clade, but not in the others, suggesting an important difference between the clades. (Table 3).
- DNA binding, transferase activity, ion binding, anion binding, and hydrolase activity were very common in CTGSer1, but not present in CTGSer2 (Table 4).
- The abundance of kinases and DNA-binding proteins suggest that CTG reassignment may play a role in regulatory functions. Protein kinases modify other proteins by inserting phosphate groups, which change the activity of the modified protein, while DNA-binding often regulate gene expression.

Further Research

- Continuing the project, a larger sample size of 125 species will be analyzed to obtain more accurate results.
- Focus on the the correlation between "CTG-ness" and regulatory functions.

Acknowledgements

Natalie Zaba thanks the Biology Department of Clark University for the support through the Summer Undergraduate Research Program (2020). This research was supported by US Department of Energy Grant: 0000238431 (D. Hibbett, PI)

References

- Bernard A Dujon, Edward J Louis, Genome Diversity and Evolution in the Budding Yeasts (Saccharomycotina). *Genetics*, Volume 206, Issue 2, 1 June 2017, Pages 717–750, <https://doi.org/10.1534/genetics.116.199216>
- Shen, Xing Xing et al. "Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum." *Cell* 175.6 (2018): 1533-1545.e20. *Cell*. Web.
- Cohen A, Nelson H, Nelson N. Metal-Ion Transporters—From Yeast to Human Diseases. In: *Madame Curie Bioscience Database* [Internet]. Austin (TX): Landes Bioscience; 2000-2013. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK6091/>
- Gerwien, Franziska et al. "Metals in fungal virulence." *FEMS microbiology reviews* vol. 42,1 (2018): fux050. doi:10.1093/femsre/fux050
- Santos, Manuel A S et al. "The genetic code of the fungal CTG clade." *Comptes rendus biologies* vol. 334,8-9 (2011): 607-11. doi:10.1016/j.crvi.2011.05.008
- Sárkány, Zsuzsa et al. "Ser or Leu: structural snapshots of mistranslation in *Candida albicans*." *Frontiers in molecular biosciences* vol. 1 27. 19 Dec. 2014, doi:10.3389/fmolb.2014.00027