Identifying Interacting Partners of Ribosomal Proteins

A THESIS

Submitted by

Akhilesh Kesavan

BS14B003

for the partial fulfillment of the requirements for the award of the degree of

DUAL DEGREE (B.TECH & M.TECH)

in

BIOLOGICAL SCIENCES

DEPARTMENT OF BIOTECHNOLOGY

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

June 2020

THESIS CERTIFICATE

This is to certify that the thesis titled **Identifying Interacting Partners of Ribosomal Proteins** submitted by **Akhilesh Kesavan (BS14B003)**, to the Indian Institute of Technology, Madras, for the award of degree of **Dual Degree (B.TECH & M.TECH) in Biological Sciences**, is a bona fide record of the research work done by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Himanshu Sinha
Research guide
Associate Professor
Dept. of Biotechnology
IIT Madras, 600 036

ACKNOWLEDGEMENTS

I would like to thank my guide Prof. Himanshu Sinha Sinha, department of biotechnology, IIT
Madras for providing guidance, assistance and the opportunity to work on this project. I would also like to thank the Initiative for Biological Systems Engineering (IBSE) for providing me a platform to present my work as well as exposure to the scientific community during my time working on the project.
I would also like to thank the members of our lab, the Systems Genetics Lab; Purnima for the motivation and drive to actually do my work; Vishwa for his immensely vast global knowledge and vocabulary; Bala for teaching me that the time of the day should not affect your work; Ashley for being a friend I could discuss things I normally can't with others; Ramya and Aditya for being the entertaining person that she always is; Nikitha, Nayana, Indu and Srijith for providing a positive research environment

Table of Contents

1. Introduction 6

1.1. What are ribosomes? 6

1.2. The motivation of the study 7

1.3. Objectives 8

1.4. Why Saccharomyces cerevisiae? 8

2. Identification of Naturally Occurring Ribosomal Protein Variants 9

2.1. Introduction 9

2.2. Database 10

3. Procedure 10

3.1. Results 13

3.2. Discussion 13

4. Generation of RP allele clones using Gateway cloning 14

4.1. Introduction 14

4.1.1. What is gateway cloning 14

4.2. Methodology 16

4.3. Results 16

4.4. Discussion 16

5. Future Work 17

5.1. Introduction 17

5.2. Why Yeast 2 Hybrid 17

5.3. Possible outcomes and hypotheses 17

6. Highlights of the work 18

7. References 19

Identifying Interacting Partners of Ribosomal Proteins

Introduction

What are ribosomes?

Protein synthesis is a vital part of a cell's housekeeping function and takes up a major portion of the cell's GTP pool[1,2]. Being such an energy-intensive process, the protein synthesis is carried out by specific macromolecular machinery called ribosomes and assisted by the transfer RNA (tRNA). The ribosome is an ancient protein making machine made up of ribosomal proteins (RP) and ribosomal RNA (rRNA). These ribosomal proteins are traditionally considered to be highly conserved across species and have very low evolution rates

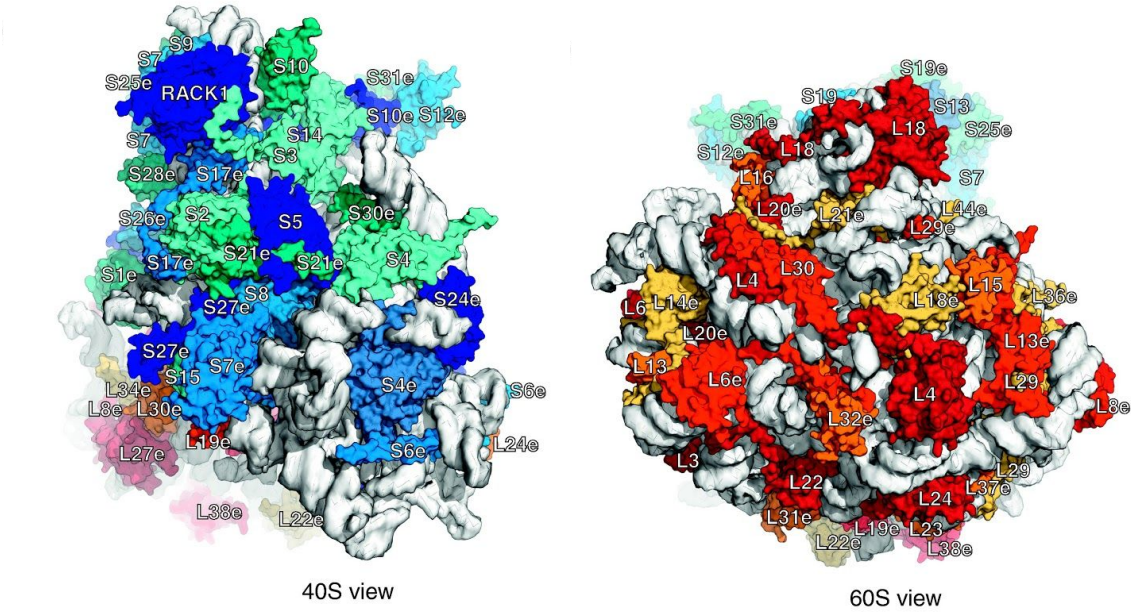


Figure 1: Crystal structures of the two subunits[3]

The motivation of the study

In an analysis conducted by the lab using data published in a previous study, it was observed that when strains of yeast whose ribosomal proteins were deleted were grown in different conditions, they show a huge phenotypic variance[4].

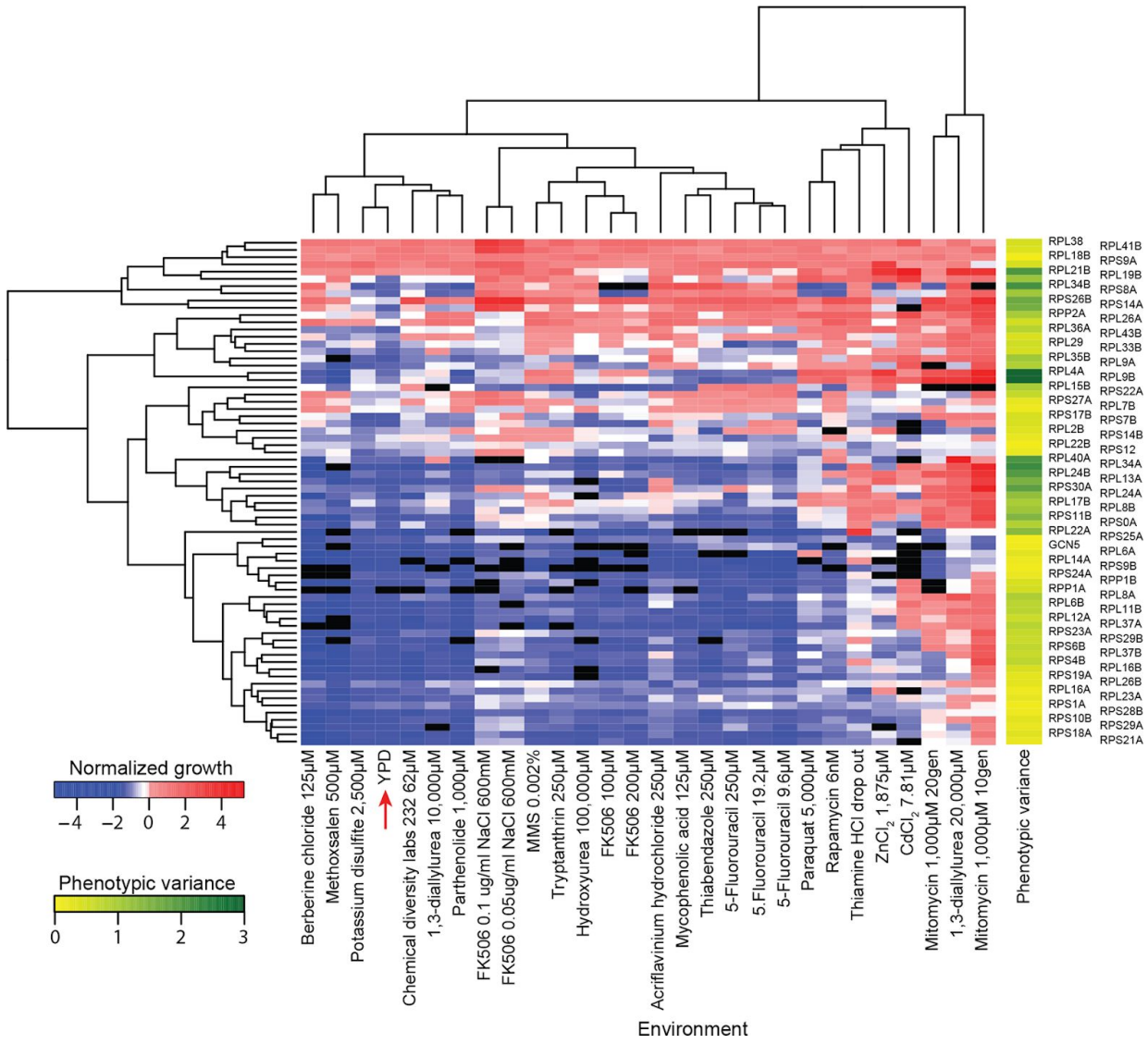


Figure 2: Phenotypic variability of yeast RP genes: Growth of deletions of RP genes (black dots) in rich medium, YPD (x-axis) versus an oxidizing agent, Cadmium chloride (CdCl2) (y-axis)[4]

Recent studies also provide evidence of some extra ribosomal activity of the ribosomal proteins like cell development[5–9], ribosome assembly[10], [10–12], selection of mRNA for translation[13], cell signaling[14] as well as possible involvement in diseases like Diamond-Blackfan anemia[14,15] to name a few. All these evidences pose an intriguing question: How do the ribosomal proteins exhibit their extra ribosomal functions?

Objectives

- Search the existing *Saccharomyces* Genome Database (SGD) and identifying in naturally existing non-synonymous variant alleles of RP genes across all 38 strains in the *Saccharomyces* Genome Resequencing Project database (SGRP)
- Isolate and amplify the genes of interest
- Clone the amplified genes of interest into pDONR223 vectors to generate a library of pENTR vectors
- Conduct yeast 2 hybrid studies using the RP genes as baits and the 5000 yeast proteins as preys and generate the list of allele-specific interacting partners for all RPs

Why *Saccharomyces cerevisiae*?

An analysis of the growth of a deletion collection of S.cerevisiae showed that deletion strains of RP genes have the highest phenotypic variability in growth across several stress environments compared to growth in standard rich medium

Identification of Naturally Occurring Ribosomal Protein Variants

Introduction

There are 37 identified strains of *Saccharomyces cerevisiae[16]*. These strains, over the years, have grown in different environments and accumulated different mutations. A small portion of these mutations is present in the ribosomal proteins. Since these mutations appear naturally in different strains, there is a reasonable possibility that these variable alleles confer some benefit to the stain with respect to its native environment over other strains. While both synonymous and non-synonymous variants are known to have an effect of the proteome, synonymous variants don't change the final ribosomal protein, thereby making its effect indirect, limited to the synthesis of the ribosomal proteins. Non-synonymous variants, however, provide a more viable opportunity for the ribosomal protein to affect the proteome actively by potentially altering key interaction of the ribosomal protein with other proteins in the cell

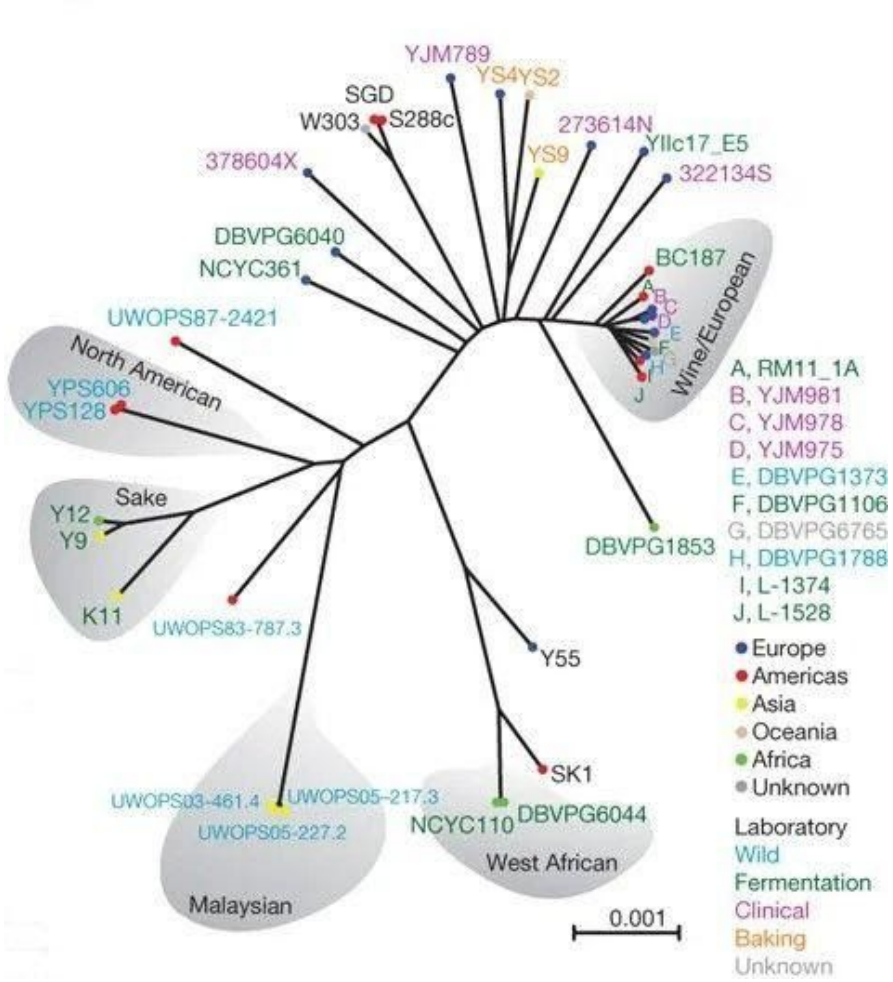


Figure 3: Neighbour-joining tree of all S. cerevisiae strains and their places of origin[16]

Database

The *Saccharomyces* Genome Database (SGD)[17] is a community resource for studying the budding yeast *Saccharomyces cerevisiae.* The SGD project provides a piece of encyclopedic information on the yeast genome and its genes, proteins and encoded features. Along with these, peer-reviewed literature of experimental results of functions and interactions is also available. All this data is provided through Locus Summary pages, a powerful query engine, and a rich genome browser. The SGD resource provides the gold standard for functional description of budding yeast, and a platform from which to investigate related genes and pathways in higher organisms.

The *Saccharomyces* Genome Resequencing Project (SGRP)[7] is a collaborative project to explore the genetic and phenotypic diversity of *Saccharomyces.* The goal of the project was to advance our understanding of genetic variation and evolution by analyzing sequences from multiple strains of the two *Saccharomyces* species, *S.cerevisiae* and *S.paradoxus*.The project sequenced haploids of 37 *cerevisiae* strains and 27 *paradoxus* strains. The sequence data has been aligned to the respective reference genome sequences. We use the blast server supplemented by this data to obtain the RP gene sequences across the various strains.

Translation of the genomic sequences is done using the online resource, EMBOSS Transeq[18]. For alignment of the various sequences, we used the online resource, Clustal Omega by EMBL[19]

Procedure

1. The sequence of the gene of interest is obtained from SGD. In case of the presence of intronic regions, the coding sequence is taken i.e. only exons
2. The sequence is BLASTed in the SGRP blast server to obtain the gene sequence in all the S.cerevisiae strains. If the gene is present in the non-sense strand of the DNA, the reverse complement is taken
3. The sequences are translated using Transeq. In case the sequence doesn't start from the beginning of the gene, the translation frame is adjusted accordingly
4. The translated sequences are aligned using the online multiple sequence alignment tool, Clustal Omega and the screened for variant amino acids
5. The variant amino acid along with the corresponding variant nucleotide, their positions wrt the start codon, and strains exhibiting the variant are listed

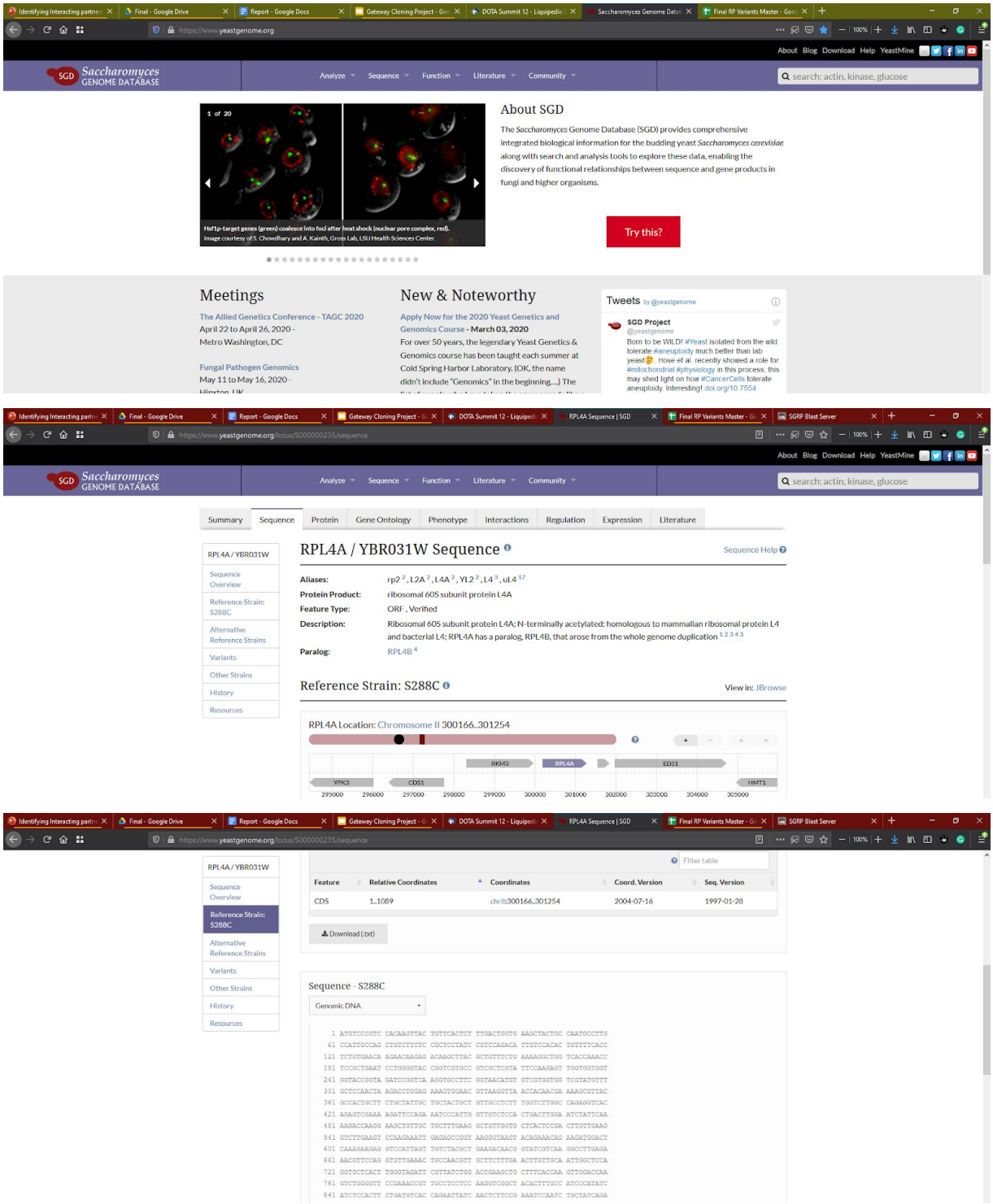
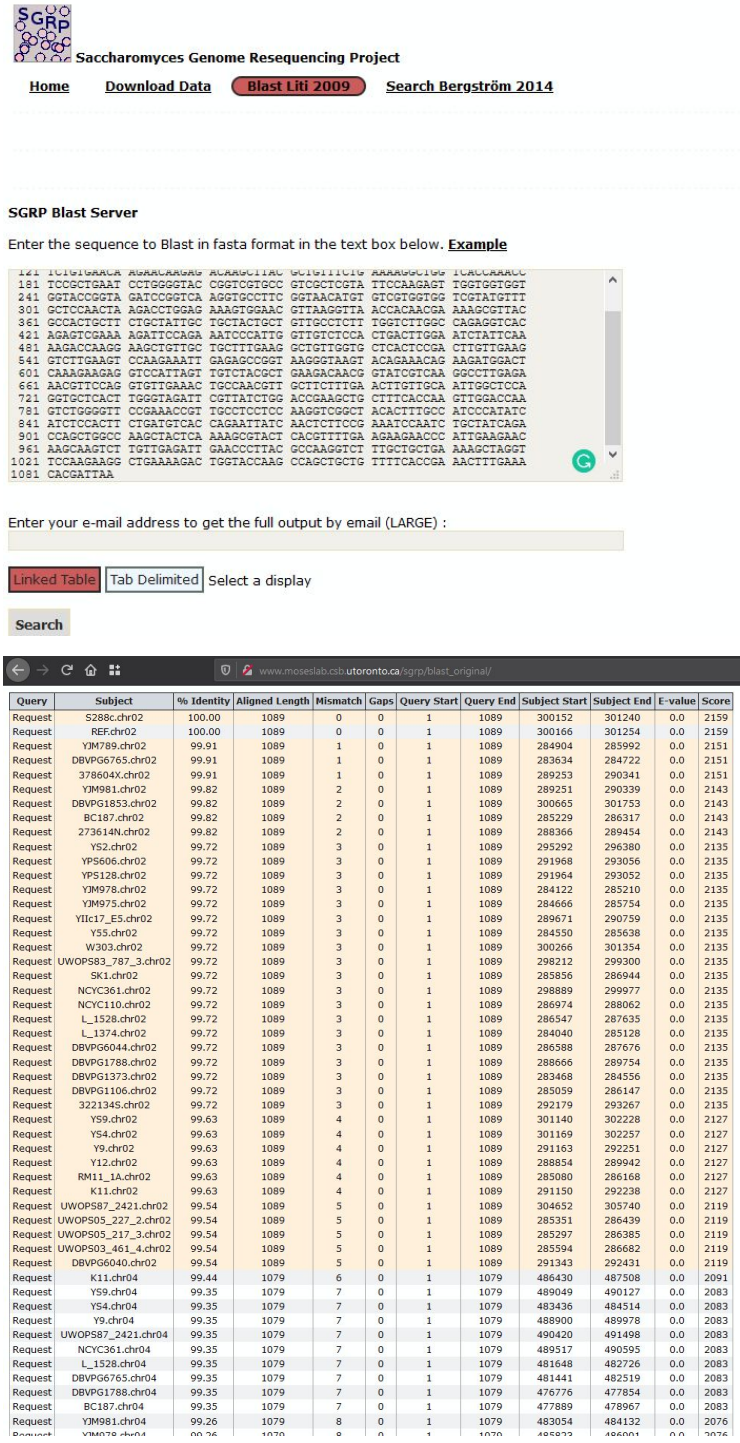


Figure 4: Screenshots from the SGD website for the RPL4A query[17]



Query	Subject	% Identity	Aligned Length	Mismatch	Gaps	Query Start	Query End	Subject Start	Subject End	E-value	Score
Request	S288c.chr02	100.00	1089	0	0	1	1089	300152	301240	0.0	2159
Request	REF.chr02	100.00	1089	0	0	1	1089	300166	301254	0.0	2159
Request	YJM789.chr02	99.91	1089	1	0	1	1089	284904	285992	0.0	2151
Request	DBVPG6765.chr02	99.91	1089	1	0	1	1089	283634	284722	0.0	2151
Request	378604X.chr02	99.91	1089	1	0	1	1089	289253	290341	0.0	2151
Request	YJM981.chr02	99.82	1089	2	0	1	1089	289251	290339	0.0	2143
Request	DBVPG1853.chr02	99.82	1089	2	0	1	1089	300665	301753	0.0	2143
Request	BC187.chr02	99.82	1089	2	0	1	1089	285229	286317	0.0	2143
Request	273614N.chr02	99.82	1089	2	0	1	1089	288366	289454	0.0	2143
Request	YS2.chr02	99.72	1089	3	0	1	1089	295292	296380	0.0	2135
Request	YPS606.chr02	99.72	1089	3	0	1	1089	291968	293056	0.0	2135
Request	YPS128.chr02	99.72	1089	3	0	1	1089	291964	293052	0.0	2135
Request	YJM978.chr02	99.72	1089	3	0	1	1089	284122	285210	0.0	2135
Request	YJM975.chr02	99.72	1089	3	0	1	1089	284666	285754	0.0	2135
Request	YIIc17_E5.chr02	99.72	1089	3	0	1	1089	289671	290759	0.0	2135
Request	Y55.chr02	99.72	1089	3	0	1	1089	284550	285638	0.0	2135
Request	W303.chr02	99.72	1089	3	0	1	1089	300266	301354	0.0	2135
Request	UWOPS83_787_3.chr02	99.72	1089	3	0	1	1089	298212	299300	0.0	2135
Request	SK1.chr02	99.72	1089	3	0	1	1089	285856	286944	0.0	2135
Request	NCYC361.chr02	99.72	1089	3	0	1	1089	298889	299977	0.0	2135
Request	NCYC110.chr02	99.72	1089	3	0	1	1089	286974	288062	0.0	2135
Request	L_1528.chr02	99.72	1089	3	0	1	1089	286547	287635	0.0	2135
Request	L_1374.chr02	99.72	1089	3	0	1	1089	284040	285128	0.0	2135
Request	DBVPG6044.chr02	99.72	1089	3	0	1	1089	286588	287676	0.0	2135
Request	DBVPG1788.chr02	99.72	1089	3	0	1	1089	288666	289754	0.0	2135
Request	DBVPG1373.chr02	99.72	1089	3	0	1	1089	283468	284556	0.0	2135
Request	DBVPG1106.chr02	99.72	1089	3	0	1	1089	285059	286147	0.0	2135
Request	322134S.chr02	99.72	1089	3	0	1	1089	292179	293267	0.0	2135
Request	YS9.chr02	99.63	1089	4	0	1	1089	301140	302228	0.0	2127
Request	YS4.chr02	99.63	1089	4	0	1	1089	301169	302257	0.0	2127
Request	Y9.chr02	99.63	1089	4	0	1	1089	291163	292251	0.0	2127
Request	Y12.chr02	99.63	1089	4	0	1	1089	288854	289942	0.0	2127
Request	RM11_1A.chr02	99.63	1089	4	0	1	1089	285080	286168	0.0	2127
Request	K11.chr02	99.63	1089	4	0	1	1089	291150	292238	0.0	2127
Request	UWOPS87_2421.chr02	99.54	1089	5	0	1	1089	304652	305740	0.0	2119
Request	UWOPS05_227_2.chr02	99.54	1089	5	0	1	1089	285351	286439	0.0	2119
Request	UWOPS05_217_3.chr02	99.54	1089	5	0	1	1089	285297	286385	0.0	2119
Request	UWOPS03_461_4.chr02	99.54	1089	5	0	1	1089	285594	286682	0.0	2119
Request	DBVPG6040.chr02	99.54	1089	5	0	1	1089	291343	292431	0.0	2119
Request	K11.chr04	99.44	1079	6	0	1	1079	486430	487508	0.0	2091
Request	YS9.chr04	99.35	1079	7	0	1	1079	489049	490127	0.0	2083
Request	YS4.chr04	99.35	1079	7	0	1	1079	483436	484514	0.0	2083
Request	Y9.chr04	99.35	1079	7	0	1	1079	488900	489978	0.0	2083
Request	UWOPS87_2421.chr04	99.35	1079	7	0	1	1079	490420	491498	0.0	2083
Request	NCYC361.chr04	99.35	1079	7	0	1	1079	489517	490595	0.0	2083
Request	L_1528.chr04	99.35	1079	7	0	1	1079	481648	482726	0.0	2083
Request	DBVPG6765.chr04	99.35	1079	7	0	1	1079	481441	482519	0.0	2083
Request	DBVPG1788.chr04	99.35	1079	7	0	1	1079	476776	477854	0.0	2083
Request	BC187.chr04	99.35	1079	7	0	1	1079	477889	478967	0.0	2083
Request	YJM981.chr04	99.26	1079	8	0	1	1079	483054	484132	0.0	2076

Figure 5: Screenshot of the BLAST search of the RPL4A sequence on the SGRP server[20]

Results

Of the 132 RP genes present in S.cerevisiae, 53 show at least one non-synonymous variant in at least one strain. A total of 81 variants were observed

Discussion

The existence of the non-synonymous variants in what is considered to be a highly conserved and essential macromolecule further supports our hypothesis that RPs are not just involved in the translational pathway but posses some extra ribosomal function as well

Generation of RP allele clones using Gateway cloning

Introduction

In order to study the effect of the variant RPs on the cell, multiple approaches can be used. One possible way is to generate allele-specific populations for each variant and compare the growth in different conditions. While this approach is viable for in-depth studies of an RP, it is not advisable for a large scale screen of the effect. To this effect, a different approach is considered.
In this approach, the variant alleles are expressed in a system and screened for all its interacting partners in the entire yeast genome. This process is more scalable to the number of variants observed and provides an understanding of the effect of the variant at a whole-cell level
Keeping this in mind, strains are chosen such that for a given RP gene there exists only one non-synonymous SNP difference between them. The genes are then extracted and cloned thereby generating two copies of the gene that, when translated differ by only one amino acid.

What is gateway cloning

Gateway cloning is a novel cloning technique developed by Invitrogen, where the gene of interest is inserted into the plasmid using recombination rather than the traditional restriction enzyme based ligation. This is achieved by flanking the gene of interest with specific sequences called att-B sites that can recombine with a complementary site on the plasmid, called att-P site in a reaction termed BP reaction. The technique was developed inspired by the way lambda phages insert and extract their genomic DNA from the host during the course of an infection. Its genome is inserted into the host using an enzyme, integrase, and an integration host factor. The reverse reaction called LR reaction is catalyzed by excisionase along with the integrase and IHF.

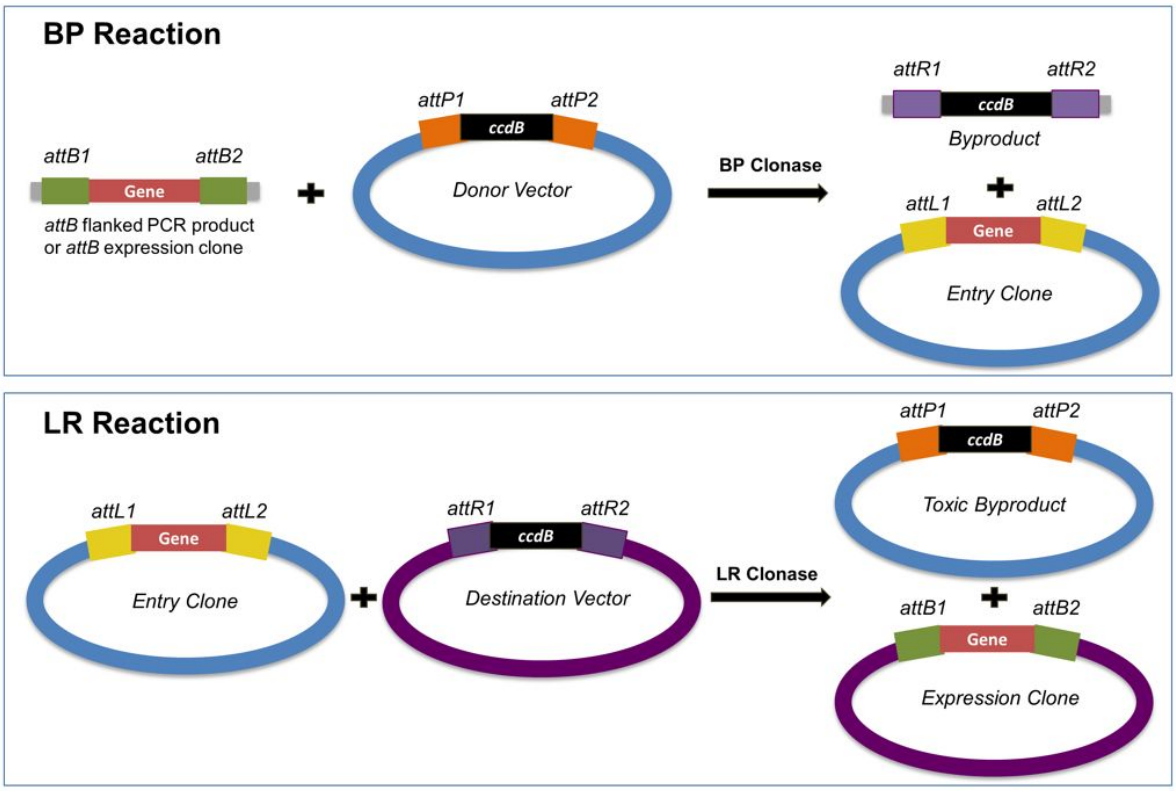


Figure 6: Gateway Cloning [21]

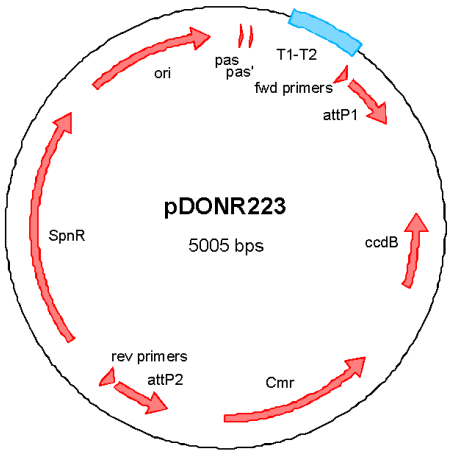


Figure 7: The plasmid map of pDONR223

Methodology

1. The strains of S.cerevisiae containing the variant allele in the gene of interest is cultured overnight and the genomic DNA is extracted
2. The gene of interest is amplified by PCR using primers that contain the attB sites flanking the nucleotide specific regions such that the amplified product is the gene of interest flanked by the attB sites
3. The amplified PCR product is then purified using 30% PEG 8000/ 30mM MgCl2 solution to remove primer dimers and other DNA
4. The purified product is then incubated with the pDONR vector and the BP Clonase™ II enzyme mix at 25℃ overnight (BP Reaction)
5. Proteinase K solution is added to terminate the reaction
6. The cloned vector is verified by gel electrophoresis
7. The BP reaction is then transformed into dh5α cells by heat shock
8. The transformed cells are grown SOC complete media to ensure better growth of the transformants
9. The transformed cells are cultured overnight and stored as glycerol stocks at -80℃

Results

The strains of S.Cerevisiae are selected such that they have a single SNP difference in the gene of interest. Therefore all the variants of each gene of interest are cloned such that they differ by one SNP

Discussion

Since the variants of a gene of interest differ by one SNP, we can study the effect of the variant allele individually. This lets us establish the effect each variant has in further studies

Future Work

Introduction

Following the creation of the variant clones, we aim to establish the interacting partners of each of the genes of interest and study the effect the variants have on them. Since we aim to screen a large number of variants across the entire proteome of S.cerevisae, we opt for the Yeast 2 Hybrid (Y2H) system for the interaction studies.

In a Yeast 2 Hybrid system, the host cell is modified by replacing a native transcription activator, like Gal4 is replaced with two synthetic hybrid proteins; one containing the N-terminal DNA binding domain of Gal 4 fused to one of our cloned variants (called bait), and the other containing the C-terminal transcription activating domain fused to one of the 5000 proteins present in the yeast genome (called prey). In a normal cell, Gal4 binds to the DNA and recruits the transcription factors in the presence of galactose media. We attempt the replicate this activity with the hybrid proteins; if the bait and prey interact the complex formed can replicate the role of the native GAL4 protein leading to transcription and measurable □-galactosidase activity. If they do not interact then no □-galactosidase activity is observed as a consequence of no transcription.

Why Yeast 2 Hybrid

While there are several techniques available to study protein interactions, it is incredibly hard to scale them to a large number of samples. The Yeast 2 Hybrid system enables us to perform a large scale study using a library of the yeast genome simultaneously. Being an in-vivo system it is better suited to replicate the conditions within the cell making the model closer to the actual yeast cell. The Yeast 2 hybrid system also takes care of all post-transcriptional and translational modifications that the protein would normally undergo. Additionally, we do not need to make proteins and antibodies of high purity which is laborious as well as expensive.

Possible outcomes and hypotheses

The Yeast 2 Hybrid studies will generate a network of genes that our protein of interest interacts with. The variant can potentially alter this network by removing and/ or adding new genes. This difference in the set of genes should be caused entirely by the variant

allele. This gives us an idea of what pathways our ribosomal protein is potentially involved in and what sort of extra ribosomal function it possesses

Highlights of the work

The key features of this work are mentioned below

1. This work builds on upcoming hypotheses that support extra ribosomal functions for the ribosomal proteins
2. Though thought to be highly conserved, variant alleles have been identified in the ribosomal proteins
3. A novel high-efficiency approach to cloning is employed in the form of gateway cloning

References

1. Kafri M, Metzl-Raz E, Jona G, Barkai N. The Cost of Protein Production. Cell Reports. 2016. pp. 22–31. doi:10.1016/j.celrep.2015.12.015

2. Thomson E, Ferreira-Cerca S, Hurt E. Eukaryotic ribosome biogenesis at a glance. Journal of Cell Science. 2013. pp. 4815–4821. doi:10.1242/jcs.111948

3. Jenner L, Melnikov S, Garreau de Loubresse N, Ben-Shem A, Iskakova M, Urzhumtsev A, et al. Crystal structure of the 80S yeast ribosome. Curr Opin Struct Biol. 2012;22: 759–767.

4. Yadav A, Radhakrishnan A, Panda A, Singh A, Sinha H, Bhanot G. The Modular Adaptive Ribosome. doi:10.1101/068361

5. Strittmatter AW, Fischer C, Kleinschmidt M, Braus GH. FLO11 mediated filamentous growth of the yeast Saccharomyces cerevisiae depends on the expression of the ribosomal RPS26 genes. Mol Genet Genomics. 2006;276: 113–125.

6. Mao-De L, Jing X. Ribosomal Proteins and Colorectal Cancer. Current Genomics. 2007. pp. 43–49. doi:10.2174/138920207780076938

7. Danilova N, Sakamoto KM, Lin S. Ribosomal protein S19 deficiency in zebrafish leads to developmental abnormalities and defective erythropoiesis through activation of p53 protein family. Blood. 2008. pp. 5228–5237. doi:10.1182/blood-2008-01-132290

8. Byrne ME. A role for the ribosome in development. Trends Plant Sci. 2009;14: 512–519.

9. Szakonyi D, Byrne ME. Ribosomal protein L27a is required for growth and patterning in Arabidopsis thaliana. The Plant Journal. 2011. pp. 269–281. doi:10.1111/j.1365-313x.2010.04422.x

10. Fernández-Pevida A, Rodríguez-Galán O, Díaz-Quintana A, Kressler D, de la Cruz J. Yeast Ribosomal Protein L40 Assembles Late into Precursor 60 S Ribosomes and Is Required for Their Cytoplasmic Maturation. Journal of Biological Chemistry. 2012. pp. 38390–38407. doi:10.1074/jbc.m112.400564

11. Martín-Marcos P, Hinnebusch AG, Tamame M. Ribosomal Protein L33 Is Required for Ribosome Biogenesis, Subunit Joining, and Repression of GCN4 Translation. Molecular and Cellular Biology. 2007. pp. 5968–5985. doi:10.1128/mcb.00019-07

12. Nissan TA. 60S pre-ribosome formation viewed from assembly in the nucleolus until

export to the cytoplasm. The EMBO Journal. 2002. pp. 5539–5547. doi:10.1093/emboj/cdf547

13. Bauer JW, Brandl C, Haubenreisser O, Wimmer B, Weber M, Karl T, et al. Specialized Yeast Ribosomes: A Customized Tool for Selective mRNA Translation. PLoS ONE. 2013. p. e67609. doi:10.1371/journal.pone.0067609

14. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, et al. Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol. 1999;17: 676–682.

15. Raiser DM, Narla A, Ebert BL. The emerging importance of ribosomal dysfunction in the pathogenesis of hematologic disorders. Leukemia & Lymphoma. 2014. pp. 491–500. doi:10.3109/10428194.2013.812786

16. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, et al. Population genomics of domestic and wild yeasts. Nature. 2009;458: 337–341.

17. Saccharomyces Genome Database | SGD. [cited 7 Jun 2020]. Available: https://www.yeastgenome.org/

18. EMBL-EBI. EMBOSS Transeq < Sequence Translation Sites < EMBL-EBI. [cited 7 Jun 2020]. Available: https://www.ebi.ac.uk/Tools/st/emboss_transeq/

19. EMBL-EBI. Clustal Omega < Multiple Sequence Alignment < EMBL-EBI. [cited 7 Jun 2020]. Available: https://www.ebi.ac.uk/Tools/msa/clustalo/

20. SGRP Blast Server. [cited 7 Jun 2020]. Available: http://www.moseslab.csb.utoronto.ca/sgrp/blast_original/

21. Soriano M. Plasmids 101: Gateway Cloning. [cited 7 Jun 2020]. Available: https://blog.addgene.org/plasmids-101-gateway-cloning