Understanding the dynamics of novel transcripts in *Saccharomyces cerevisiae*

Alejandra Gonzalez
Alejandro Agudelo
Ting Chun, Huang

Table of Contents

1. Introduction

2. Results and Discussion

3. Methods
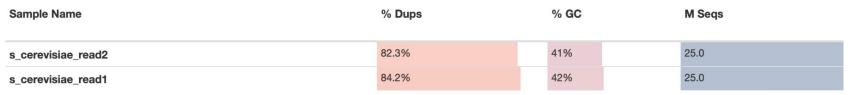
4. Conclusions

5. Bibliography

1. INTRODUCTION

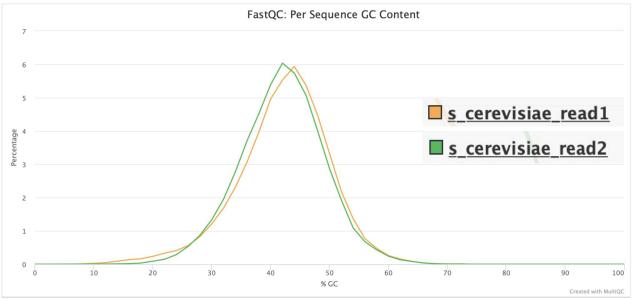S.cerevisiae

De novo genes

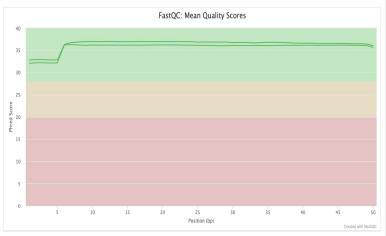Evolutionary pathways

Gene annotations vs transcriptomics

2. RESULTS & DISCUSSION

RNA-Seq reads have high-quality scores

Sample Name	% Dups	% GC	M Seqs
s_cerevisiae_read2	82.3%	41%	25.0
s_cerevisiae_read1	84.2%	42%	25.0

FastQC: Per Sequence GC Content

2. RESULTS & DISCUSSION

RNA-Seq reads have high-quality scores
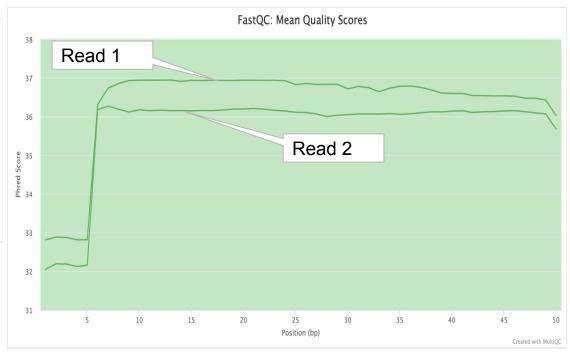

FastQC: Mean Quality Scores

MultiQC: Summarize analysis results for multiple tools and samples in a single report
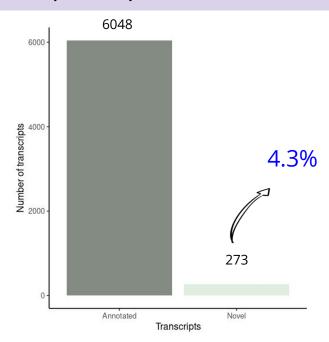*Philip Ewels, Måns Magnusson, Sverker Lundin and Max Käller*
Bioinformatics (2016)
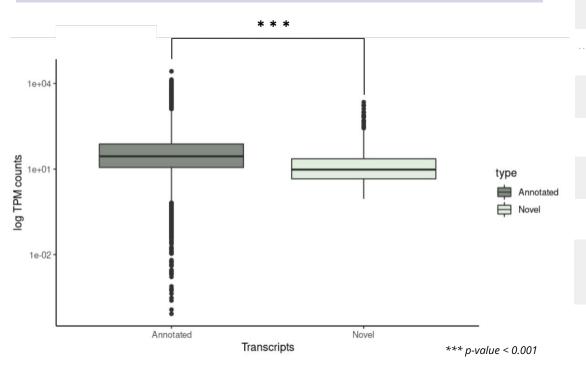doi: 10.1093/bioinformatics/btw354
PMID: 27312411

FastQC: Mean Quality Scores

Read 1

Read 2

2. RESULTS & DISCUSSION

Identification of annotated and novel transcripts

Quantification of transcript expression   (TPM) counts

2. RESULTS & DISCUSSION

Annotated transcripts are more expressed than novel

*** p-value < 0.001

Average counts (TPM):
- **Annotated**   161.96 +/- 744.79
- **Novel**   72 +/- 249.5

de novo genes (≠ strand)

↓

Novel transcripts

↓
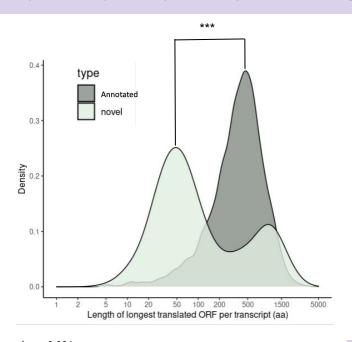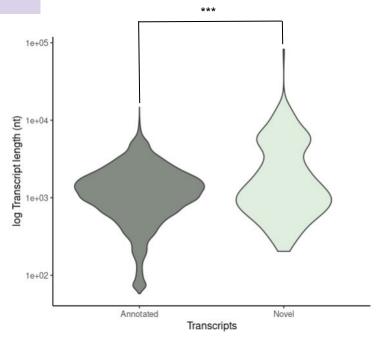
Transcriptional relationship between annotated and novel transcripts

2. RESULTS & DISCUSSION

Identification of novel putative protein-coding transcripts

NOVEL TRANSCRIPTS

- Non-protein coding transcripts

- small proteins

- transcriptional noise

*** p-value < 0.001

2. RESULTS & DISCUSSION

Identification of novel putative protein-coding transcripts

*** p-value < 0.001
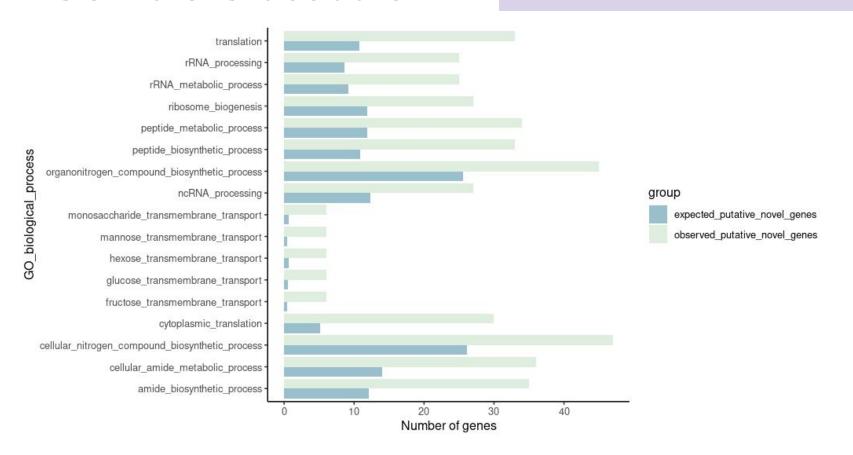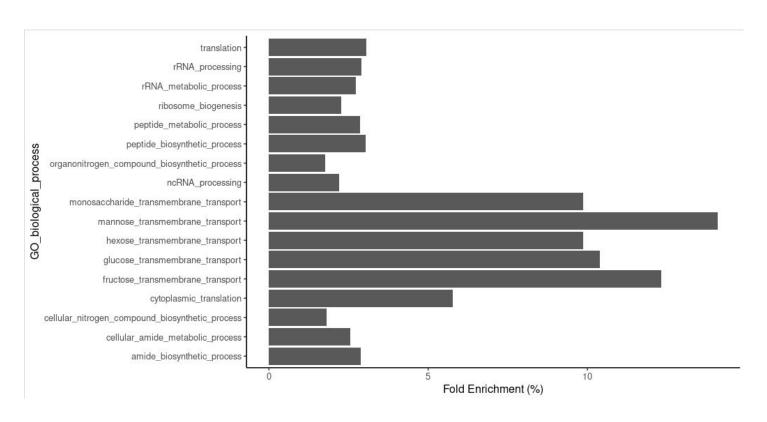
Novel transcripts have shorter sequences than annotated ones

2. RESULTS & DISCUSSION

2. RESULTS & DISCUSSION

3. METHODS

Raw data → **Quality control**

Raw data → **Alignment** (Hisat 2.1.0) → **Sam file** → **Convert to bam format** (Samtools 1.10) → **Assemble transcripts** (Stringtie v2.1.1)

Convert to bam format → **Visualization of reads** (IGV 2.11.4)

Assemble transcripts → **Visualization of reads** (IGV 2.11.4)

Assemble transcripts (Stringtie v2.1.1) → **Obtain sequences in fasta format** (Bedtools v2.27.1) → **Identify ORFs** → **Find homology** (Blastn 2.9.0) → **Enrichment Analysis** (geneontology.org)

3. METHODS

4. CONCLUSIONS

- Gene annotations and transcriptomics data gave 273 posible novel transcripts.

- 206 of them may translate to a protein with a function.

- There is overexpresion of 13 biological processes in the novel transcripts which may show us a posible evolutionary pathway of S.cerevisiae.

- Our study has limitations and further studies are needed.

6. BIBLIOGRAPHY

- Draghici, S. (2018, August 22). *ADVAITA*. Retrieved November 28, 2021, from ADVAITA: https://advaitabio.com/faq-items/understanding-gene-ontology/

- Geneontology. (2021, November 28). *Geneontology*. Retrieved from Geneontology: http://geneontology.org/

- Stanford University. (2021, November 28). *Yeast genome*. Retrieved from Saccharomyces Genome Database (SGD): https://www.yeastgenome.org/

- William R. Blevins, J. R.-O.-M.-C. (2019, March 13). Frequent birth of de novo genes in the compact yeast genome. *bioRxiv*. doi:https://doi.org/10.1101/575837