

Saccharomyces mikatae

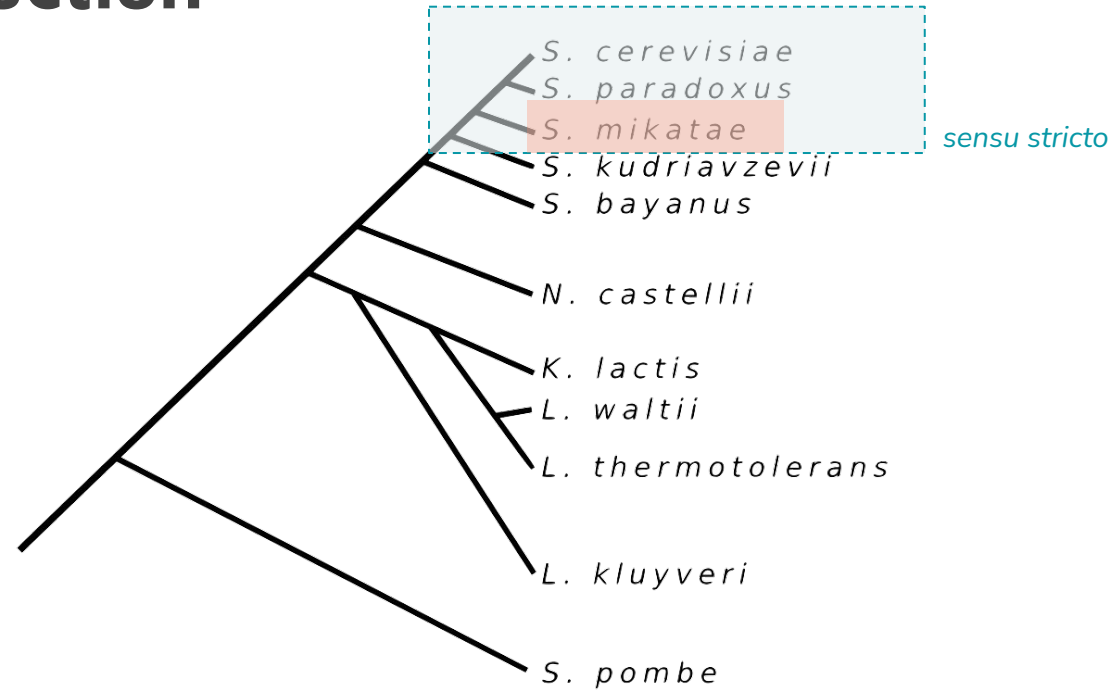
A genome characterization project

Maria Artigues, Pablo Choquehuanca and Marta Huertas





Introduction

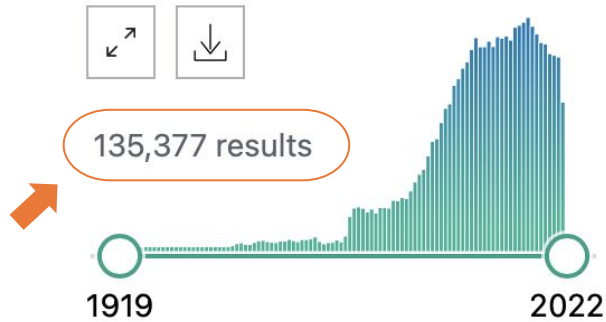


Uncovering de novo gene birth in yeast using deep transcriptomics. Blevins et. al. 2021



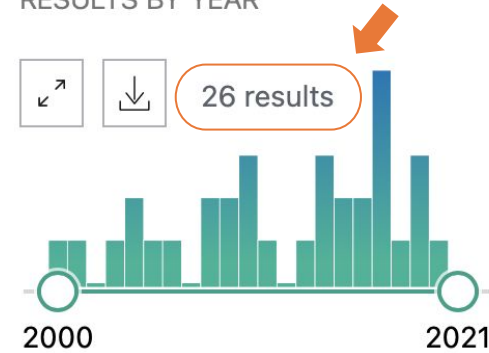
Introduction

RESULTS BY YEAR



Search term:
Saccharomyces cerevisiae

RESULTS BY YEAR



Search term:
Saccharomyces mikatae

Introduction

Are there statistical differences between novel and known genes in *S. mikatae*?

First results showed high amount of novel genes

Are most non-annotated genes **novel** or is it just a **bad annotation**?

Hypothesis

Non-annotated genes are mostly novel genes and have specific properties such as being shorter, less expressed, etc.

Compare novel and known genes

Similar properties could mean non-annotated genes are not novel.
Poor annotation

Different properties could mean non-annotated genes are mostly novel.

Methodology

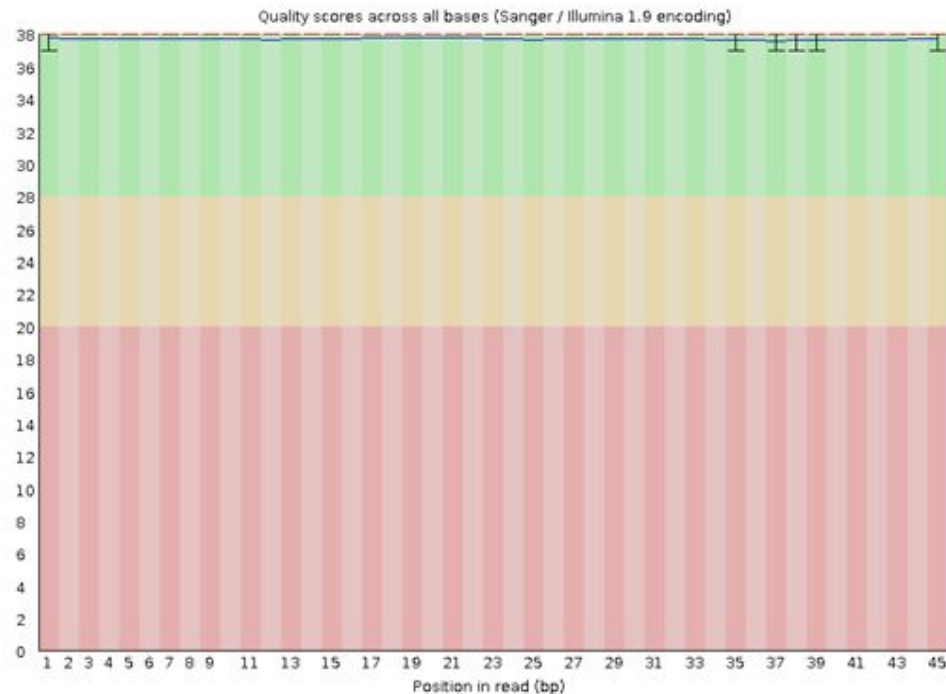
DATA

R1: s_mikatae_read1
R2: s_mikatae_read2

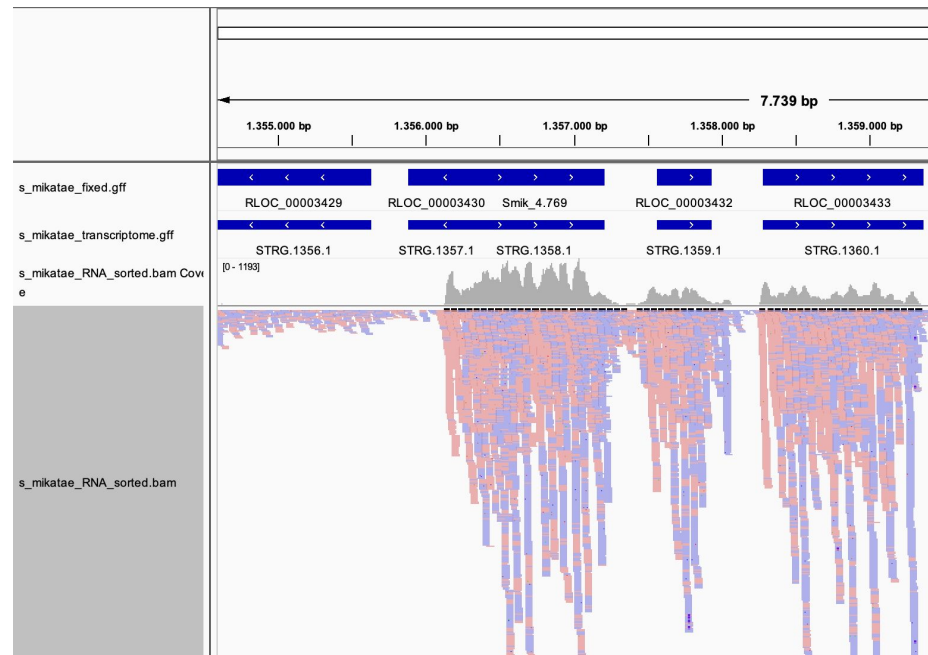
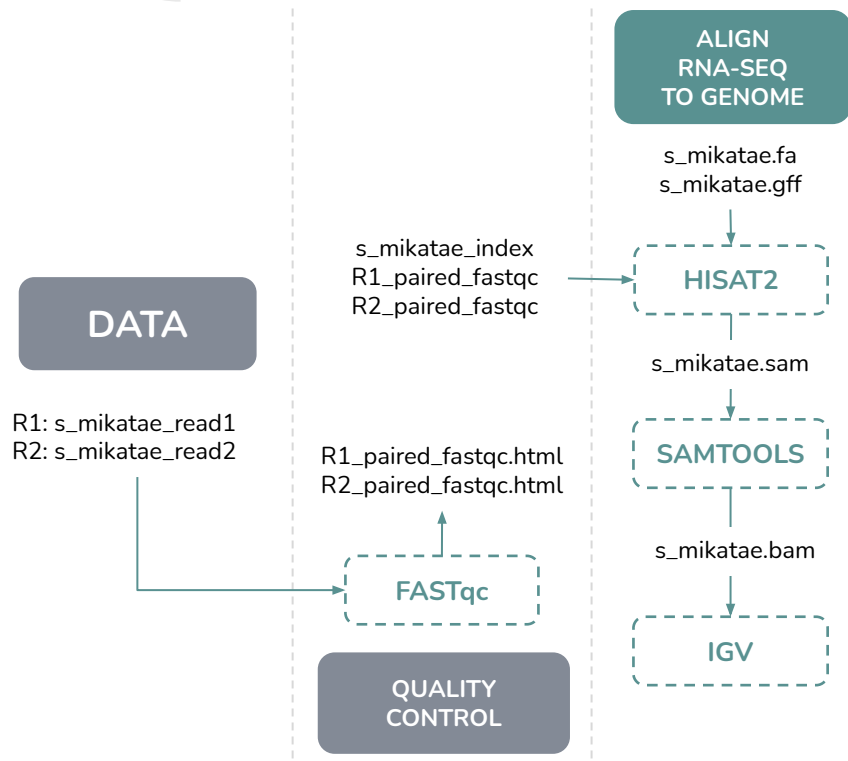
R1_paired_fastqc.html
R2_paired_fastqc.html

FASTqc

QUALITY
CONTROL

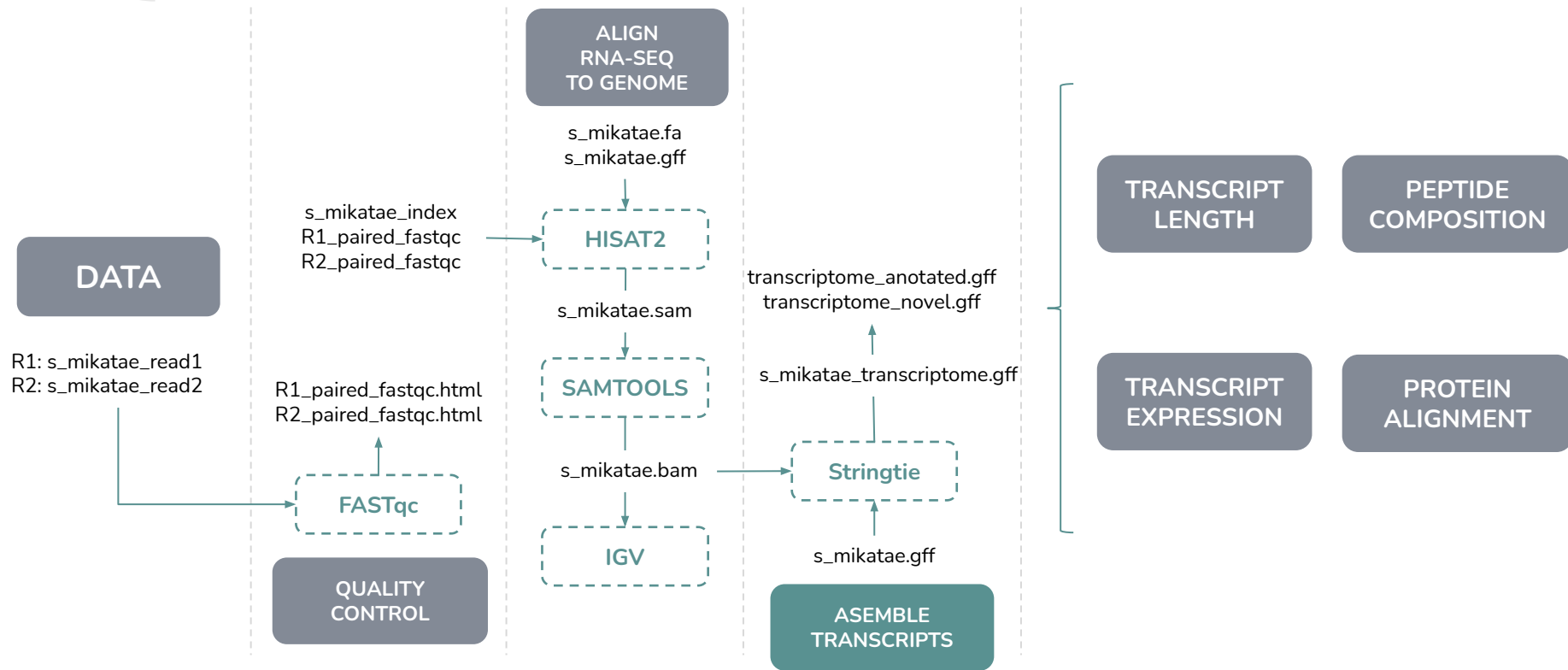


Methodology

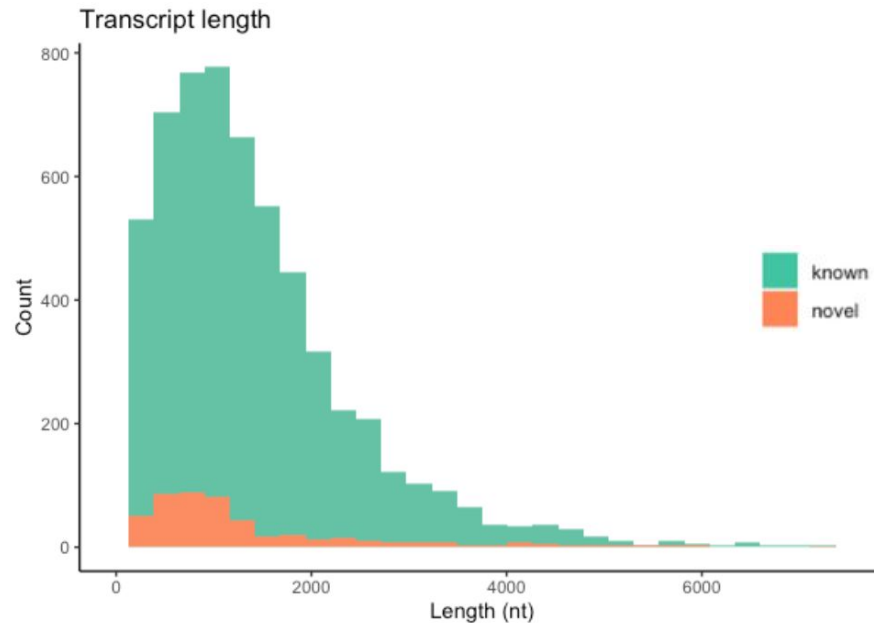


IGV screenshot

Methodology



Transcript length



Wilcoxon test: p-value = 0.2186

Known

Mean	1379
Median	1140

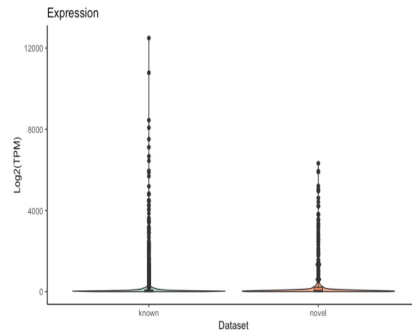
n = 5921

Novel

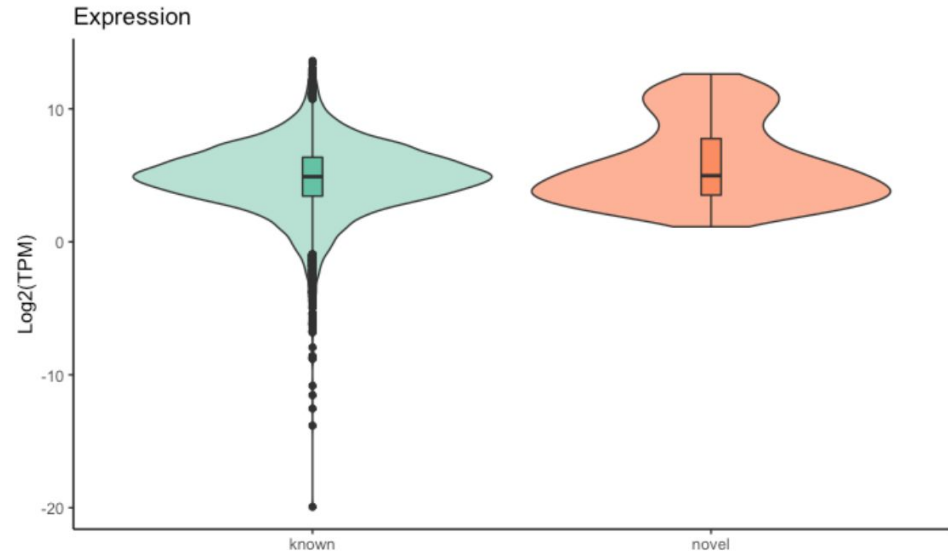
Mean	1580
Median	978

n = 499

Transcript expression



Log₂



Known

Mean 123.07

Median 29.89

n = 5921

Novel

Mean 489.52

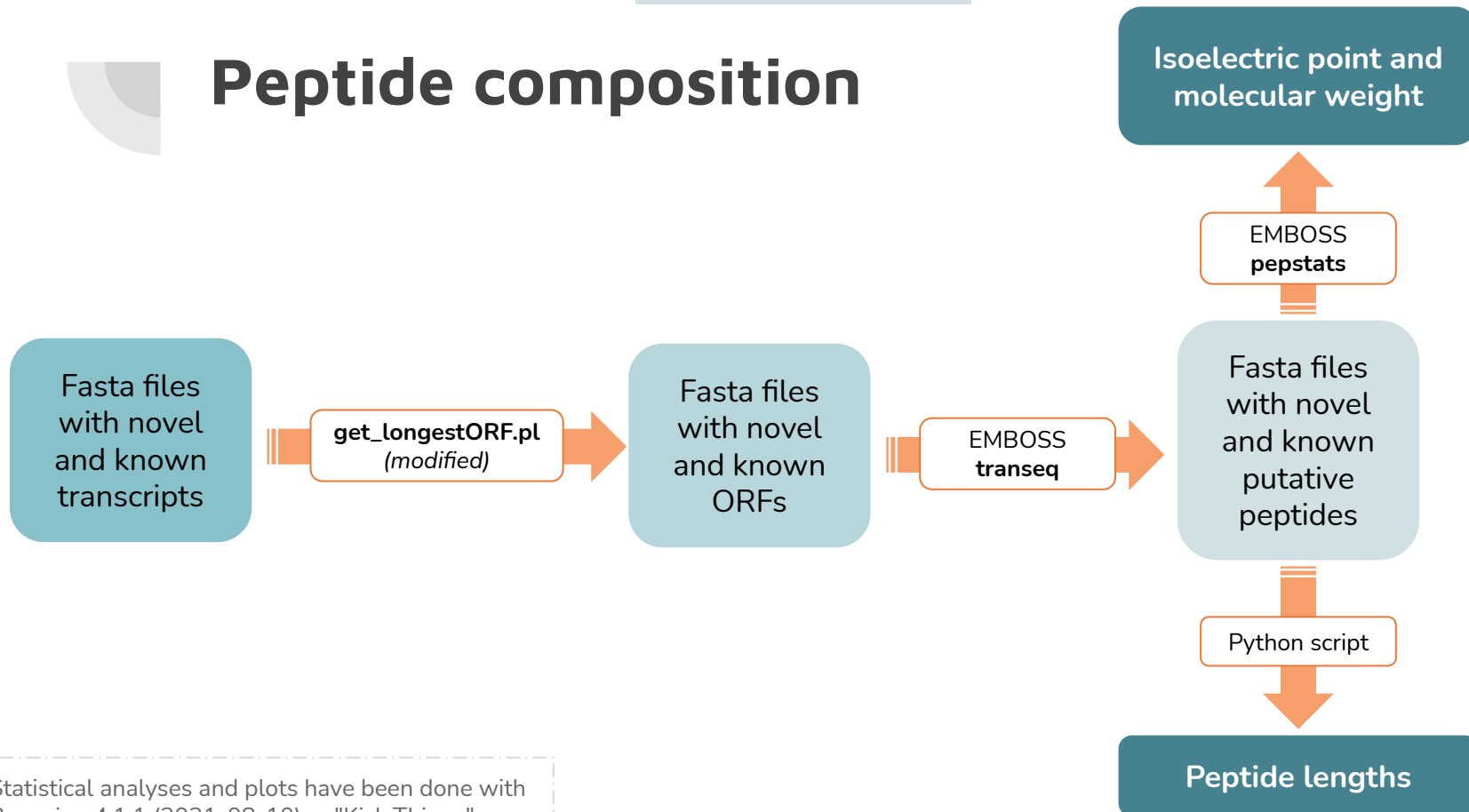
Median 31.48

n = 499

Wilcoxon test: p-value > 2.2e-16

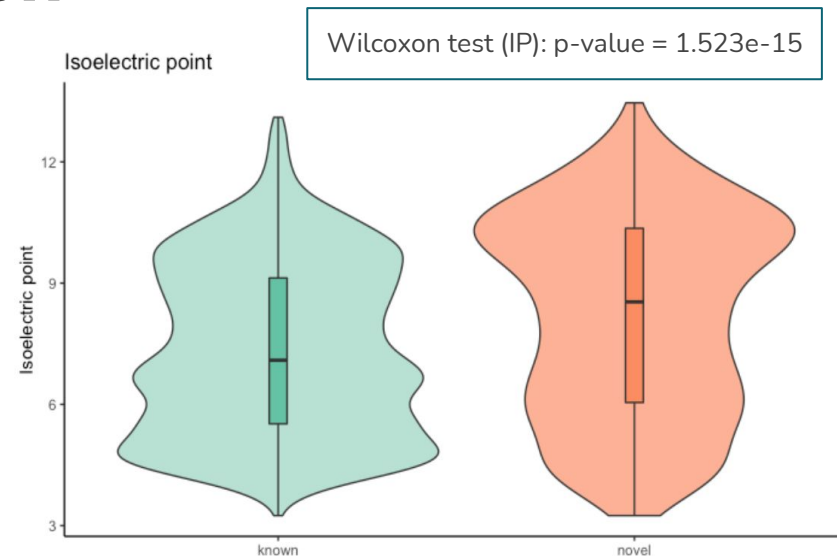
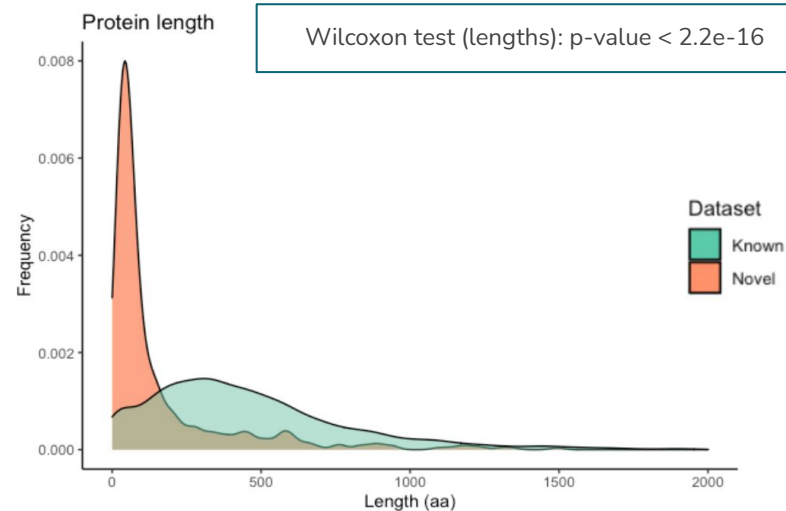


Peptide composition



Statistical analyses and plots have been done with R version 4.1.1 (2021-08-10) -- "Kick Things"

Peptide composition

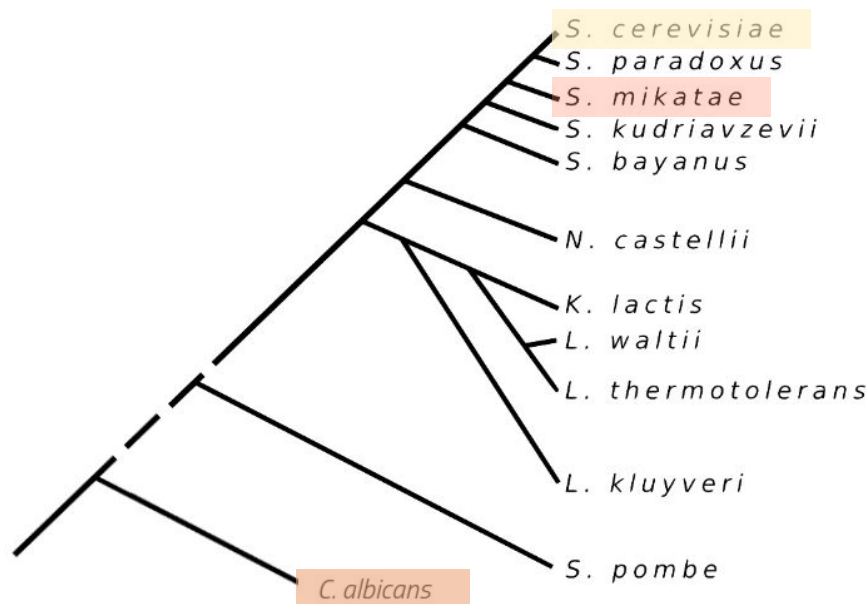


Most
common
amino acids

Known	M, T, J and L
Novel	M, T, L and J

	n	Length mean	Isoelectric point mean
Known	5810	467.3	7.34
Novel	485	163.3	8.23

Protein alignment



Modified from: Uncovering de novo gene birth in yeast using deep transcriptomics. Blevins et. al. 2021

S. mikatae peptides

blastp

Vs. *S. cerevisiae* proteome

Vs. *C. albicans* proteome

No hits with *S. cerevisiae*

Hits only with *S. cerevisiae*

Hits with *S. cerevisiae* and *C. albicans*

Length of ORFs by type

	Novel ORFs <i>S. mikatae</i>	No hits with <i>S. cerevisiae</i>	Hits only with <i>S. cerevisiae</i>	Hits with <i>S. cerevisiae</i> and <i>C. albicans</i>
n	486	226	60	200
Mean	1535.1	983.7	987.4	2322.4
Median	970.5	759.0	758.0	1397.0

Wilcoxon test

No hits *cerevisiae* Vs. Only *cerevisiae*

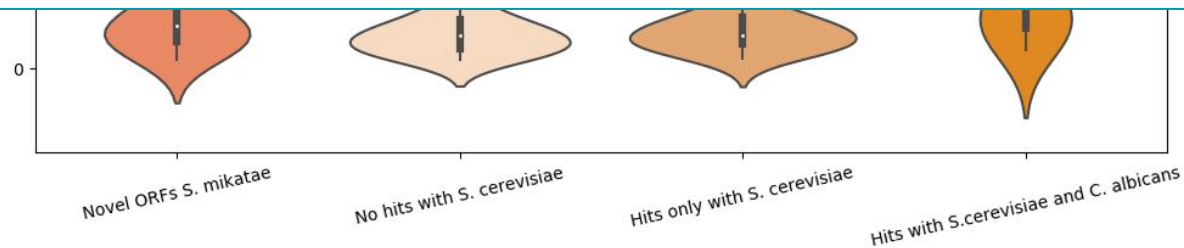
p-value = 0.3014

No hits *cerevisiae* Vs. Hits *cerevisiae* and *albicans*

p-value < 2.2e-16

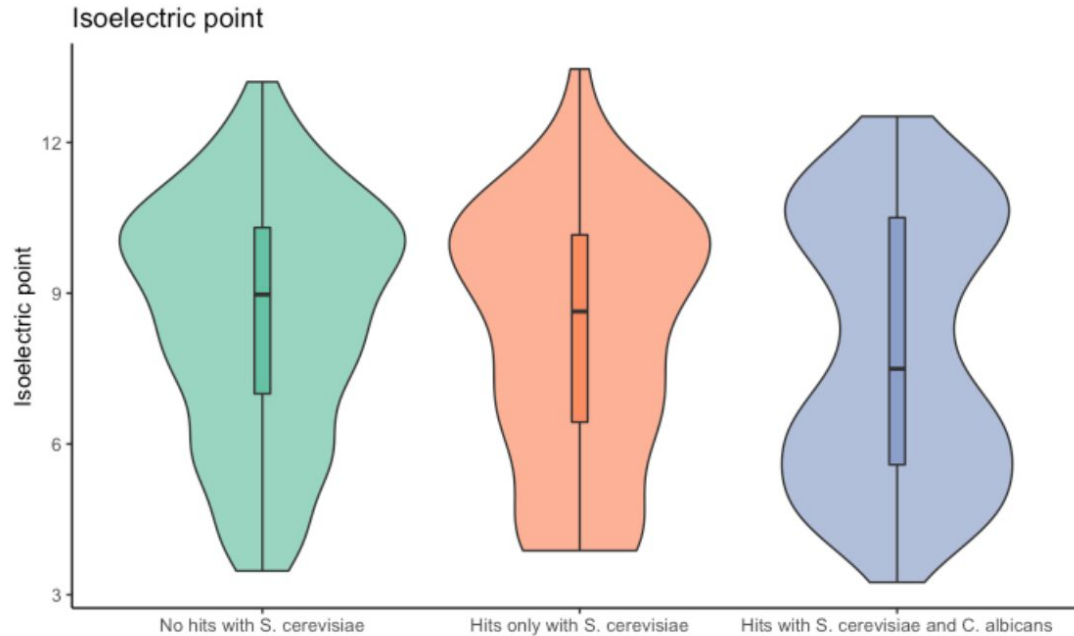
Only *cerevisiae* Vs. Hits *cerevisiae* and *albicans*

p-value = 3.508e-10





Isoelectric point





Take home message



Differential characteristics between known and novel genes: length, expression and peptide composition.



'Novel' annotation as a mix of putative novel and poorly annotated genes.



Lengths difference between putative novel and poorly annotated.



Improve methodology to produce more robust results.



Bibliography

Albà M, Castresana J. Inverse Relationship Between Evolutionary Rate and Age of Mammalian Genes. *Molecular Biology and Evolution*. 2005;22(3):598-606.

Blevins WR, Ruiz-Orera J, Messegueur X, Blasco-Moreno B, Villanueva-Cañas JL, Espinar L, et al. Uncovering de novo gene birth in yeast using deep transcriptomics. *Nature communications*. 2021;12(1):1-3.

Naumov GI, James SA, Naumova ES, Louis EJ, Roberts IN. Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. *International journal of systematic and evolutionary microbiology*. 2000;50(5):1931-42.



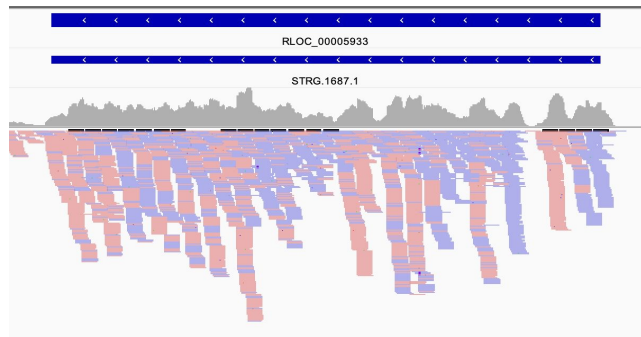
Code availability

<https://github.com/marta-huertasp/PGB-S.mikatae>

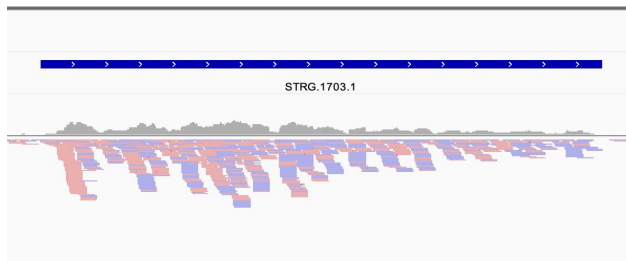


Annexes

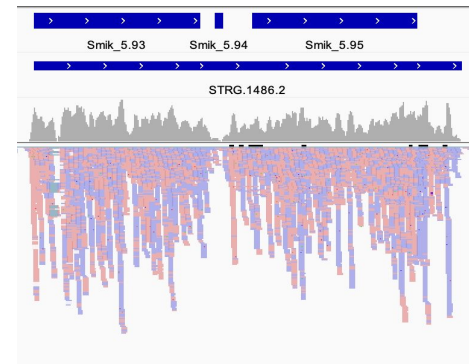
Known



Novel



Stringtie error



IGV screenshots.