Identifying outlier yeast strains from a pool of a thousand strains in maltose growth conditions

UCLA B.I.G Summer Bruins in Genomics

CINDY VO[1,2], Chantle Swichkow[2], Heriberto Marquez[2], Leslie Alamo[2], Joshua Bloom[2], Leonid Kruglyak[2]

[1] B.I.G. Summer Program, Institute of Quantitative and Computational Biosciences, UCLA[2] Department of Human Genetics, David Geffen School of Medicine, UCLA

UCLA QCBio

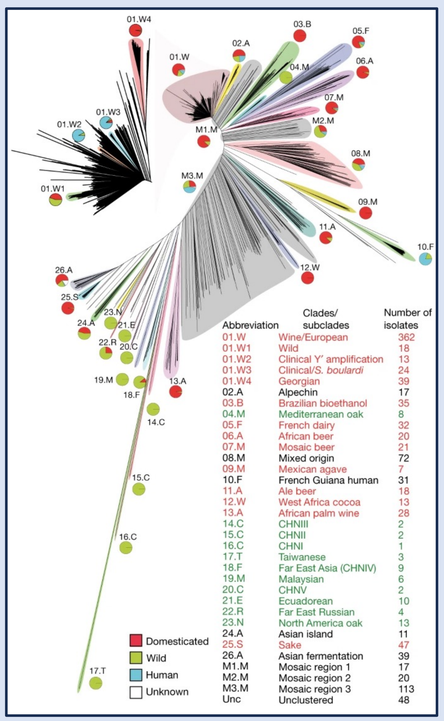
ABSTRACT

Genetic variation in yeast impacts fitness in various environments. Saccharomyces cerevisiae strains have diversified genetically due to evolution in different settings. To study the genetic differences that contribute to varying fitness, genome-wide association studies (GWAS) or meiotic recombination techniques are employed to map quantitative trait loci (QTL). The aim is to identify strains or genetic variants with enhanced fitness in maltose growth conditions from a pool of one thousand, three hundred, and eight yeast strains. Genomic DNA is extracted from each representative pool, sequenced, and analyzed, using non-negative least squares regression, to calculate genotype frequency differences before and after growth in maltose conditions. The anticipated result is the identification of genetic variations from strains better adapted to maltose conditions. From this information, a pipeline is created to identify strains with higher fitness and outlier strains in different conditions for genetic mapping.

BACKGROUND

- *Saccharomyces cerevisiae:* a yeast species that plays a pivotal role in its ability to carry out various fermentation processes.
- *S. cerevisiae*'s genetic diversity is also influenced by its long history of domestication and continuous evolution alongside human activities like brewing and baking.
- To better understand and harness this diversity, we use quantitative trait loci (QTL) mapping.
  - One such trait of interest is maltose metabolism, crucial for brewing and baking processes.
  - By applying QTL mapping, we can identify specific regions in the yeast's genome associated with maltose utilization efficiency, ultimately leading to the development of strains with enhanced maltose fermenting capabilities.
  - Some strains cannot grow in maltose. We should be able to see this signature in our phenotypic assay.
- This knowledge of genetic diversity and QTL mapping in *S. cerevisiae* opens up exciting avenues for tailored yeast strains that can optimize brewing and fermentation processes, ensuring the production of high-quality beverages and fueling advancements in biotechnology and food industries.

METHODOLOGY



Abbreviation	Clades/ subclades	Number of isolates
01.W	Wine/European	362
01.W1	Wild	18
01.W2	Clinical Y' amplification	13
01.W3	Clinical/*S. boulardi*	24
01.W4	Georgian	39
02.A	Alpechin	17
03.B	Brazilian bioethanol	35
04.M	Mediterranean oak	8
05.F	French dairy	32
06.A	African beer	20
07.M	Mosaic beer	21
08.M	Mixed origin	72
09.M	Mexican agave	7
10.F	French Guiana human	31
11.A	Ale beer	18
12.W	West Africa cocoa	13
13.A	African palm wine	28
14.C	CHNIII	2
15.C	CHNII	2
16.C	CHNI	1
17.T	Taiwanese	3
18.F	Far East Asia (CHNIV)	9
19.M	Malaysian	6
20.C	CHNV	2
21.E	Ecuadorean	10
22.R	Far East Russian	4
23.N	North America oak	13
24.A	Asian island	11
25.S	Sake	47
26.A	Asian fermentation	39
M1.M	Mosaic region 1	17
M2.M	Mosaic region 2	20
M3.M	Mosaic region 3	113
Unc	Unclustered	48

Figure 1: A phylogenic tree of the thousand strains that are pooled together. Peter et al., 2018. *Nature* 556, 339–344.

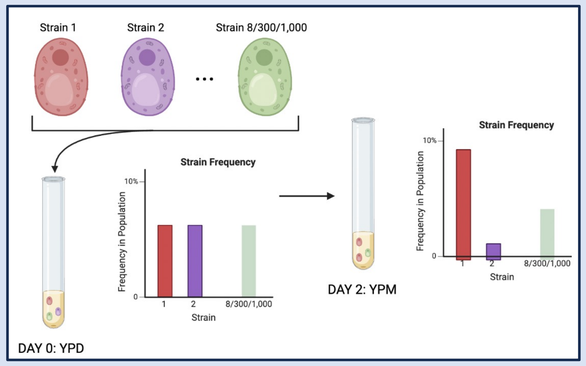


Figure 2: The experimental design to identify outlier strains in maltose growth conditions. The thousand strains are pooled together at first in YPD, a complete media used to maintain yeast culture, and is, later, transferred to YPM, a maltose selective media, for a two-day growth period. A subset of eight and three-hundred strains from the thousand collection undergo the same experimental set up.

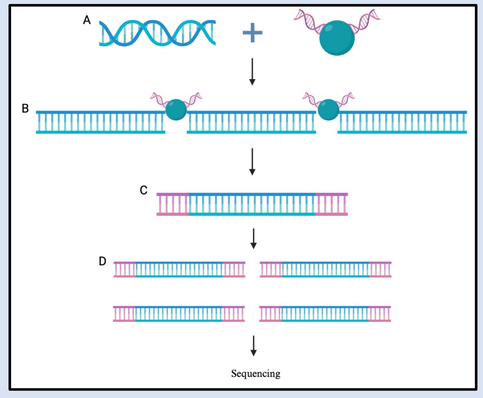


Figure 3: gDNA extraction and final library preparation process. (a) gDNA was extracted from the Day 0 and Day 2 samples and transposases added. (b) tagmentation to cleave the gDNA and add adapters. (c) gDNA fragment with adapter attached. (d) polymerase chain reaction (PCR)

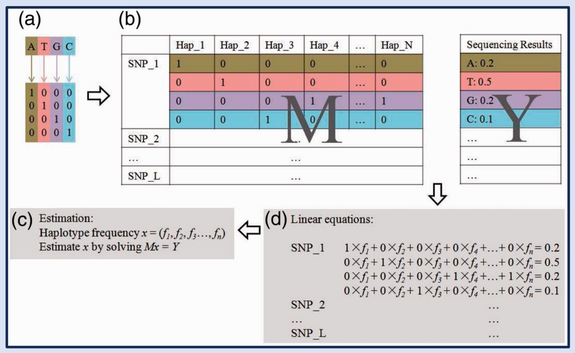


Figure 4: Non-negative least squares regression method to infer strain frequency. (a) assign nucleotide bases to binary vectors. (b) encode the haplotype database and pooled sequencing results. (c) estimating haplotype frequency. (d) linear equations are constructed for each base proportions for each SNP. Chang-Chang et al., (2015). *Bioinformatics*, Volume 31, Issue 4, February 2015, Pages 515–522

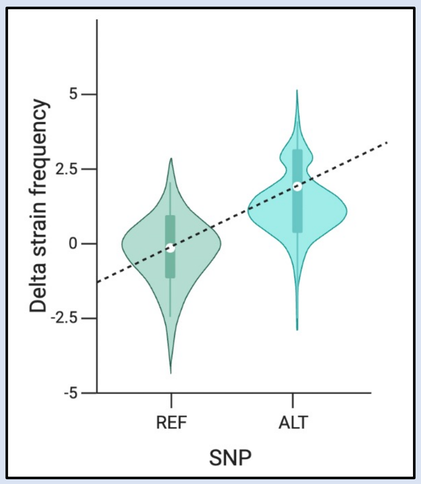


Figure 5: A genome-wide association study (GWAS) to correlate strain frequency to the inheritance of specific variants using delta strain frequency and single nucleotide polymorphism (SNP).

RESULTS

- Due to technical difficulties, we were unable to sequence and analyze the pools as described.
- The anticipated results would be an increased prevalence of strains with the *MAL* gene loci, maltose utilization genes.
- We will be able to empirically test what the maximum complexity of pooling is by identifying the percentage of strains that drop out through the various subsets of pools.
- Additionally, we can assess the sequencing depth required to adequately capture the extent of complexity by analyzing the input culture.

CONCLUSION

- The resulting pipeline unlocks the ability to quantify growth differences in complex environments, such as sourdough.
- Once growth differences are quantified, we can link the difference in genetic variants with GWAS or QTL mapping by crossing outlier strains.

ACKNOWLEDGEMENTS

Thank you to the Kruglyak Lab, particularly Chantle Swichkow, Shivani Patel, and Joshua Bloom, for their support and guidance throughout the last eight weeks. Thank you to my fellow B.I.G. program students in the Kruglyak Lab, Brooke Bowman and Alyssa Stainton, for their encouragement and companionship. Lastly, thank you to the B.I.G. Summer program for this opportunity to expand my computational knowledge and skills.

REFERENCES

[1]Chang-Chang Cao , Xiao Sun, Accurate estimation of haplotype frequency from pooled sequencing data and cost-effective identification of rare haplotype carriers by overlapping pool sequencing, *Bioinformatics*, Volume 31, Issue 4, February 2015, Pages 515–522, https://doi.org/10.1093/bioinformatics/btu670

[2]Peter, J., De Chiara, M., Friedrich, A. *et al.* Genome evolution across 1,011 *Saccharomyces* cerevisiae isolates. *Nature* 556, 339–344 (2018). https://doi.org/10.1038/s41586-018-0030-5