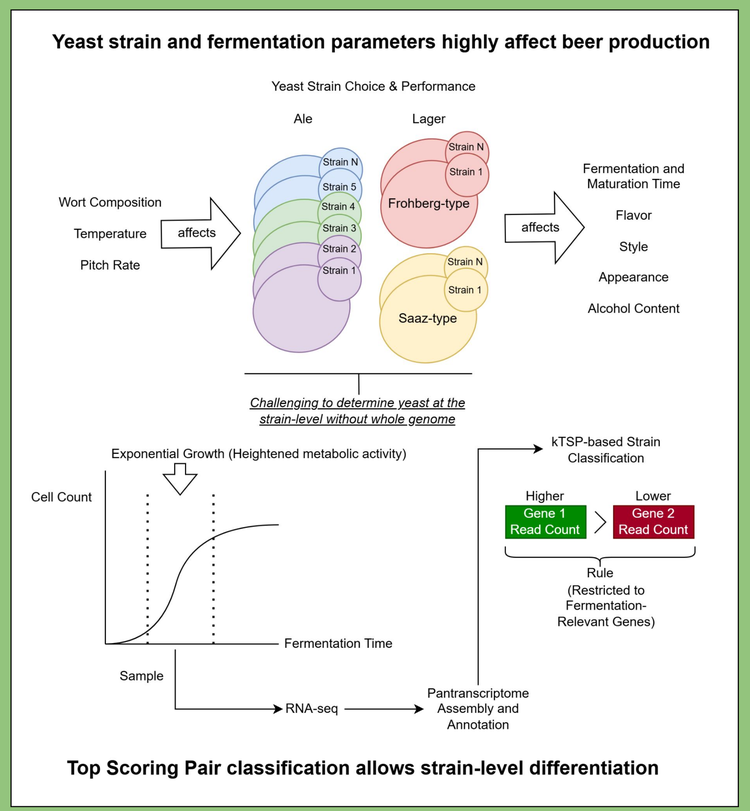
Classifying Beer Yeast Strains Using Comparative Transcriptomics

Patrick Rynkiewicz, MS
David VanHoute, PhD
Spencer Richman, MS

Abstract

Beer fermentation represents a controlled microbial ecosystem with direct impacts on food production and flavor chemistry. *Saccharomyces* beer yeasts are well-known for their diverse sensory contributions to beer, with hundreds of available commercial strains. However, accurate and cost-effective strain-level identification of beer yeasts and their differences with respect to fermentation performance remain a challenge. Here, we present a workflow identifying beer yeast at the strain level using comparative transcriptomics and a rule-based machine learning classifier trained on fermentation-relevant gene expression. Although we trained on a small sample set, we show that this methodology displays robustness to differences in fermentation temperature and growth media.

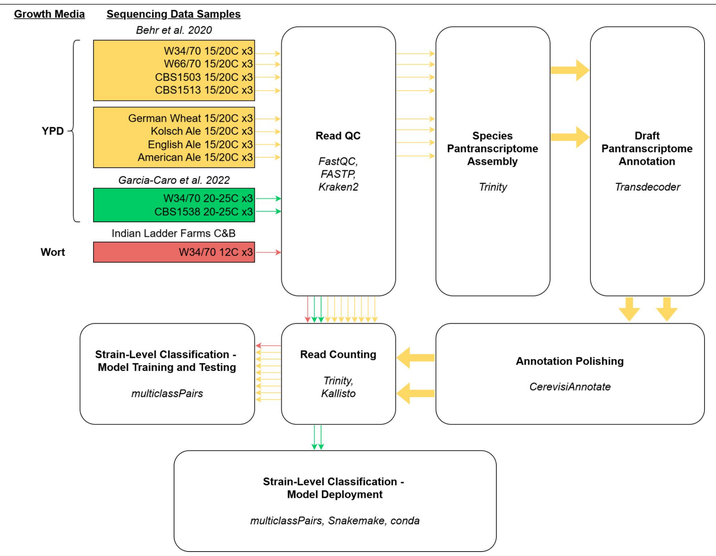
Graphical Abstract



Background

Beer represents one of humanity's oldest managed microbial ecosystems, with yeast fermentation performance being key to flavor development and microbial community control primarily through the production of ethanol and lactic acid [1,2] Commercial beer yeast strains are ultimately defined by metabolic pathways relevant to beer flavor and fermentation performance, and disambiguation of beer yeast at the strain-level using genetic methods continues to be a significant challenge [2]. Previous work has demonstrated efficacy of comparative transcriptomics methods in differentiating broadly between beer and lager yeast grown in laboratory conditions [3].

"Simple patterns in gene expression relevant to fermentation accurately classify related beer yeast at the strain-level across varying growth conditions."



Methods

Data Acquisition and Generation

- RNA-seq data was obtained from the SRA for eight yeast strains (four lager, four ale) cultured at 15°C and 20°C on YPD media. Two additional lager strain datasets were reserved as holdout test data.
- A production sample was obtained from a Märzen-style lager brewed at Indian Ladder Farms Cidery & Brewery (Altamont, NY) using Fermentis SafLager™ W-34/70 (Saaz-type) dry yeast. Yeast was harvested 48-hours post-pitch, processed, and sequenced in technical triplicate on the DNBseq platform (150bp paired-end reads).

Transcriptome Assembly and Annotation

- Following quality and contamination filtering, separate ale and lager pantranscriptomes were assembled using *Trinity* with data from Behr et al. 2020.
- Transcripts were annotated using Transdecoder and further refined with our custom CerevisiAnnotate workflow, which:
  - Disambiguated Trinity genes with competing functional annotations
  - Enhanced context where reference strain S288C annotations were utilized

Strain Classification Model

- Read counts for fermentation-relevant genes were used to train strain-specific Top Scoring Pairs models using the multiclassPairs R package.
- The classifier was validated using holdout test data including CBS1538 (Frohberg-type, not used in training) and one W34/70 sample.

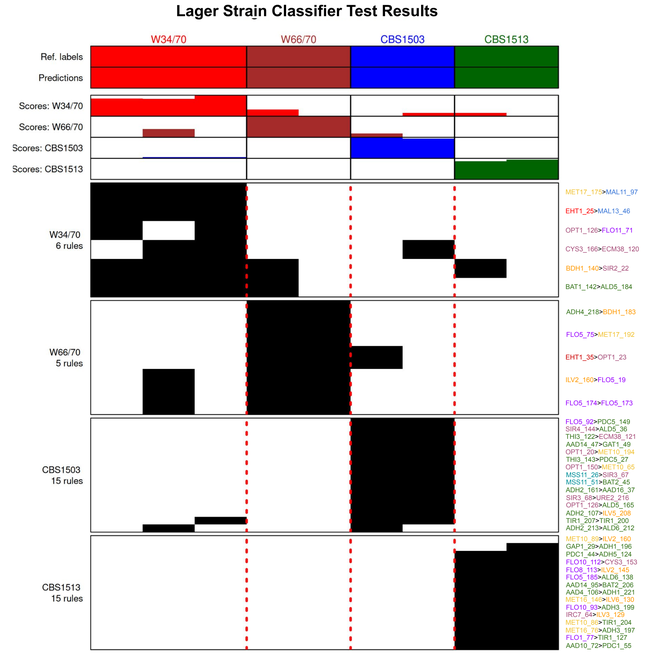
Transcriptome Assembly and Annotation

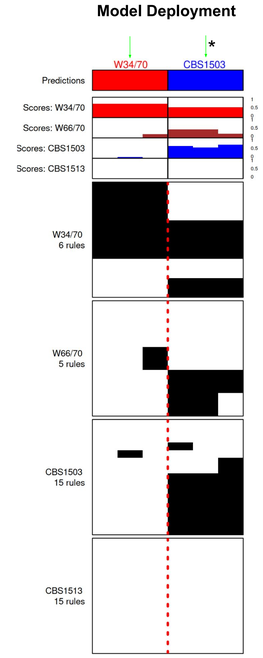
- Initial Trinity+Transdecoder assembly and draft annotation returned >20,000 sequence records with a UniProt annotation for each pantranscriptome.
- CerevisiAnnotate consolidated contradictory protein annotations into single representative records per gene based on frequency, length and alignment score.
- The majority of final annotations in both ale and lager pantranscriptomes mapped to S288C reference-strain homologs. A smaller subset retained original UniProt annotations with ordered locus names, indicating high-quality reference alignments.

Results

Pantranscriptome Annotation Metrics Before and After Refinement with CerevisiAnnotate

Pan-Transcriptome Annotation	Sequence Records with a UniProt Annotation	Unambiguous Annotations (Transdecoder)	Collapsed Annotations (CerevisiAnnotate)	Original Strain Annotations Output	S288C Homolog Annotations Output
Ale Yeasts (German Wheat, Kolsch Ale, English Ale, American Ale)	23,329	12,248	2,460	1,132	13,518
Lager Yeasts (W34/70, W66/70, CBS1503, CBS1513)	21,532	7,265	2,849	1,016	9,030





Classifier Performance

- The strain-specific classifier successfully categorized all samples using 5-15 discriminatory rules per strain in both training (not shown) and testing data (left panel).
- Visualization shows rule application across samples: black boxes indicate rules met, white boxes indicate rules not satisfied.
- Frohberg-type strain (CBS1503, CBS1513) classification relied heavily on rules involving in fusel alcohol production, flocculation, and thiol production.
- Frohberg-type strains required more discriminatory rules than Saaz-type strains (W34/70, W66/70).
- Some rules utilized alternative transcript assemblies of the same gene.
- When applied to independent samples (right panel):
  - All W34/70 production samples were correctly classified
  - CBS1538* (a Frohberg-type strain not used in training) was appropriately classified with its genetic relative CBS1503

Limitations and Future Directions

- The current study trained on a limited number of samples, and model overfitting is inevitable with the current data.
- Brewery samples were collected in technical triplicate.
- Various yeast strain families including Kveik yeast, saison yeasts, and genetically modified / hybrid lager strains are not included in the current model.
- Future work includes expanding the model to additional beer yeast strain families, collecting additional biological and technical replicates, and comparing model outputs on bioprospected / GMO strains to other yeast performance assays.

Pathway	Acetate Ester Biosynthesis	Diacetyl Processing	Ehrlich Pathway (Fusel Alcohol)	Sulfur Species Processing	Thiol Production	Maltose Utilization	Flocculation	Starch Degradation
Genes	EEB1, ATF1, ATF2, EAT1, EHT1, IAH1, SNF8	ILV2, ILV3, ILV5, ILV6, BDH1, BDH2	BAT1, BAP2, SFA1, PDC1, BAT2, ALD4, ALD5, ALD6/ALD1, ARO8, ARO9, ARO10, ARO80, PDC5, PDC6, THI3, ADH1, ADH2, ADH3, ADH4, ADH5, ADH6, ADH7/YCR105W, AAD3, AAD4, AAD6, AAD10, AAD14, AAD15, AAD16, YPL088W, TIR1, GAP1, GAT1	MET2, MET3, MET14, MET16, MET5, MET10, MET17, SSU1	IRC7, CYS3, GLN3, URE2, SIR2, SIR3, SIR4, OPT1, ECM38, STR3	MAL11, MAL12, MAL13, MAL61, MAL62, MAL63, MAL31, MAL32, MAL33, IMA1, AGT1	FLO1, FLO5, FLO8, FLO9, FLO10, FLO11	STA1, STA2, MSS11

References

1. He Y., Dong J., Yin H., Zhao Y., Chen R., Wan X., Chen P., Hou X., Liu J. and Chen L. (2014). Wort composition and its impact on the flavour-active higher alcohol and ester formation of beer – a review, *J. Inst. Brew.*, 120: 157–163. DOI: 10.1002/jib.145
2. Olaniran, A. O., Hiralal, L., Mokoena, M. P., and Pillay, B. (2017) Flavour-active volatile compounds in beer: production, regulation and control. *J. Inst. Brew.*, 123: 13–23. doi: 10.1002/jib.389.
3. Behr J, Kliche M, Geißler A, Vogel RF. Exploring the potential of comparative de novo transcriptomics to classify Saccharomyces brewing yeasts. PLoS One. 2020 Sep 23;15(9):e0238924. doi: 10.1371/journal.pone.0238924. PMID: 32966337; PMCID: PMC7510981.

Acknowledgements

Special thank-you to collaborators Dietrich Gehring and Scott Veltman, who allowed us to sample their production lager batch.

INDIAN LADDER FARMS
ALTAMONT, NY