Fantastic Budding Yeasts and How They Evolved to Be

Antonis Rokas
Department of Biological Sciences, Vanderbilt University
http://www.rokaslab.org/  @RokasLab

Art by Jacob Steenwyk
The Evolution of Budding Yeast Biodiversity

- Sequence the genomes of all ~1,100+ known budding yeast species
- Construct their definitive phylogeny and timetree & use it to revise their taxonomy
- Examine how genomic variation transforms into metabolic variation across an entire subphylum

Hittinger / Kurtzman / Groenewald / Rokas Labs; http://y1000plus.org
The Evolution of Budding Yeast Biodiversity

Hittinger lab

C. Kurtzman
taxonomy

M. Groenewald
taxonomy

Rokas lab
evolution
why yeasts?
What is a Yeast?

Yeasts: fungi that asexually reproduce by budding or fission, which results in growth that is comprised mainly of single cells

Kurtzman et al. (2011) The Yeasts, 5th Ed.

**Ascomycetes**

**Basidiomycetes**

**Rest of Fungi**

**Cryptococcus yeasts**

**Malassezia yeasts**

**Fission yeasts** (Schizosaccharomyces)

**Budding yeasts** (Saccharomyces, Candida)
Not All *Saccharomycotina* Species are Yeasts

**Diagram:**

- **Ovoid:** Starmerella bombicola
- **Triangular:** Candida boreocaroliniensis
- **Spherical:** Candida albicans, Saccharomyces cerevisiae, Hanseniaspora osmophila, Candida dubliniensis
- **Apiculate:**
- **Hyphae length**

**Measurements:**
- 0 μm
- 10 μm
- 20 μm
- 30 μm

**Additional:**
- Bacilliform
- Ellipsoidal
- Hyphae width
- Cylindrical

*Chavez et al. (2024) FEMS Yeast Res.*
Budding Yeasts Have Very Small Genomes

http://book.bionumbers.org/how-big-are-genomes/
Budding Yeasts Have Amazing Genomic Diversity

Budding Yeasts Have Genomic “Flexibility”

S. cerevisiae – S. eubayanus divergence ≈ human – chicken divergence

Budding Yeasts Vary Widely in their Diets

Nectar yeasts (Wickerhamiella / Starmerella clade)

Cactophilic yeasts

Oil producers (Lipomyces, Yarrowia)

Human pathogenic yeasts (Candida)

Plant pathogens (Eremothecium sp.)
### Strains & Metabolite Growth Data for All 1,100+ Species

#### Fermentation

<table>
<thead>
<tr>
<th>Sugar</th>
<th>Lactose</th>
<th>Raffinose</th>
<th>Trehalose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galactose</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Sucrose</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Maltose</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Growth (in Liquid Media)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inulin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sucrose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raffinose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melibiose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galactose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lactose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trehalose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maltose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melezitose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methyl-α-d-glucoside</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soluble starch</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cellobiose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salicin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>l-Sorbose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>l-Rhamnose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>l-Xylose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d-Xylose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>l-Arabinose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d-Arabinose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Budding yeasts are a fantastic system for studying the evolution of the genotype – phenotype map and for inferring principles of genome evolution in eukaryotes
Pathway Discovery using Phylogenomic Footprinting

Krause et al. (2018) PNAS
Pathway Discovery using Machine Learning

**Model:** Random Forest

Environmental, metabolic, and/or genetic features

<table>
<thead>
<tr>
<th>Species 1</th>
<th>X₁</th>
<th>X₂</th>
<th>...</th>
<th>Xₙ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>...</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Species 2</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>1</td>
</tr>
<tr>
<td>Species 3</td>
<td>0</td>
<td>1</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Species n</td>
<td>1</td>
<td>1</td>
<td>...</td>
<td>1</td>
</tr>
</tbody>
</table>

Training (90%) - Grows on galactose

Testing (10%) - Does not grow on galactose

**Feature Importance**

- Maltose
- L-Arabinose
- D-Xylose
- Cycloheximide 0.1
- GAL102
- Lactose
- Galactitol
- GAL10
- GAL7
- GAL1

**Figure 8** The reductive d-galactose degrading pathway. *Broken lines* Alternative catabolic pathways

**Fekete et al. (2004) Arch. Microbiol.**

**Harrison et al. (2023) bioRxiv**
Most of these 14 species grow on both galactose and galactitol.
Galactose Consumption and Enzymatic Activity

**Oxidoreductive Pathway**

- Absorbs light at 340 nm
- NADPH
- NADP+
- D-galactose
- Putative aldose reductase
- Galactitol

**Graphs**

- **Candida ruelliae**
  - [Galactose] (g/L) vs. Time (h)
  - Cellular density (OD_600) vs. Time (h)

- **Absorbance at 340 nm (a.u.) vs. Time (h)**

*Harrison et al. (2023) bioRxiv*
Signatures of Convergent Evolution in Cactophiliic Yeasts

Goncalves et al. (2023) bioRxiv
Leveraging AI and Phylogenomics

a. Machine learning

Phenotypic data
Genomic data

Model: Random Forest

Metabolic and/or genetic features

Species 1
Species 2
Species 3
Species n

\[ X_1 \quad X_2 \quad \ldots \quad X_n \]

Training (90%)
Testing (10%)

Cactophilic
Not cactophilic

b. Gene family evolution

c. (Relative) evolutionary rates

Species Tree
Gene Tree

Goncalves et al. (2023) bioRxiv
Specific Traits Predict Cactophily

Confusion matrix, without normalization

- True label
- Predicted label

Top 10 important features

- 37°C
- Trehalose
- Sucrose
- α-Methyl-Glucoside
- Raffinose
- K11511
- K01192
- Ribose
- 40°C
- K19783

Proportion of species

Features

Goncalves et al. (2023) bioRxiv
Convergence in Genes Involved in Cell Wall & Envelope

Goncalves et al. (2023) bioRxiv
Hanseniaspora Genomic Diversity Exceeds that of Vertebrates

Hanseniaspora Divide Rapidly, Have Bipolar Budding

https://www.youtube.com/watch?v=qFN9oZe5VlM
Genetik Universität Osnabrück
Hanseniaspora are the Dominant Yeasts on Grapes

Hanseniaspora *Lifestyle Stems from Extensive Gene Losses*

- high evolutionary rate $\iff$ loss of DNA repair genes
- fast growth $\iff$ loss of cell cycle genes
- specialized diet $\iff$ loss of metabolic genes / pathways
Hanseniaspora *Lost Many DNA Repair and Maintenance Genes*

<table>
<thead>
<tr>
<th>Outgroup ▼</th>
<th>SEL</th>
<th>FEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Present</td>
<td></td>
<td></td>
</tr>
<tr>
<td>▼ Absent</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TDP1** repairs damage from topoisomerase activity

**POL32** fills in gaps after excision

**PHR1** repairs UV damage

**MAG1** excises bases in the base-excision repair pathway

Hanseniaspora Yeasts Lost Many Cell Cycle Genes

Hanseniaspora Yeasts Lost Many Metabolic Genes

Reconstructing the Evolution of Yeast Metabolism Across the Subphylum

<table>
<thead>
<tr>
<th>Fermentation</th>
<th>Glucose</th>
<th>+ Lactose</th>
<th>–</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Galactose</td>
<td>– Raffinose</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Sucrose</td>
<td>– Trehalose</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Maltose</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Growth (in Liquid Media)</th>
<th>Glucose</th>
<th>+ d-Ribose</th>
<th>–</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inulin</td>
<td>– Methanol</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Sucrose</td>
<td>– Ethanol</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Raffinose</td>
<td>– Glycerol</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Melibiose</td>
<td>– Erythritol</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Galactose</td>
<td>– Ribitol</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Lactose</td>
<td>– Galactitol</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Trehalose</td>
<td>– d-Mannitol</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Maltose</td>
<td>– d-Glucitol</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Melezitose</td>
<td>– myo-Inositol</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Methyl-α-d-glucoside</td>
<td>– d-Lactate</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Soluble starch</td>
<td>– Succinate</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Cellobiose</td>
<td>+ Citrate</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Salicin</td>
<td>+ d-Gluconate</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>l-Sorbose</td>
<td>– d-Glucosamine</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>l-Rhamnose</td>
<td>– N-Acetyl-d-glucosamine</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>d-Xylose</td>
<td>– Hexadecane</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>L-Arabinose</td>
<td>– Nitrate</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>d-Arabinose</td>
<td>– Vitamin-free</td>
<td>–</td>
</tr>
</tbody>
</table>
Loss Exceeds Gain for 38/45 Metabolic Traits

Inferring the Metabolic Capabilities of Yeast Ancestors

BYCA (Budding Yeast Common Ancestor) was a Generalist

We infer that BYCA could grow on 28 of the 45 substrates

Probability of Trait Presence in Ancestor

*Melibiose*

*Galactose*

Widespread Loss of Traits

Widespread Loss of Traits


MYA

S. cerevisiae

BYCA

Number of traits

MYA
Widespread Loss of Traits

with so much loss, how did new metabolic traits evolve?
Horizontal Gene Transfer (HGT)
Distribution of 878 Horizontally Acquired Genes

HGT in 226 yeasts with universal code: 0.071%
HGT in 103 yeasts with non-universal code: 0.025%

W/S clade yeasts acquired the enterobactin operon from Enterobacteria – organisms from both lineages co-occur in insect guts, where iron is a growth-limiting factor.
W/S Yeasts with Functional Operon Produce Enterobactin

Blue Agar CAS Assay after 9 days of growth

Kominek, Doering, et al. (2019) Cell
“The PIs assume “built it and they will come”. While the yeast community has a long track record of capitalizing on new resources, the PIs are cautioned against arguing for innovation using “field of dreams” reasoning”

Anonymous grant reviewer (2012)
Evidence for loss and reacquisition of alcoholic fermentation in a fructophilic yeast lineage

Carla Gonçalves¹, Jennifer H Wisecaver²,³, Jacek Kominek⁴,⁵,⁶,⁷, Madalena Salema Oom¹,⁸, Maria José Leandro⁹,¹⁰, Xing-Xing Shen², Dana A Opulente⁴,⁵,⁶,⁷, Xiaofan Zhou¹¹,¹², David Peris¹,⁵,⁶,⁷,¹³, Cletus P Kurtzman¹⁴,‡, Chris Todd Hittinger⁴,⁵,⁶,⁷, Antonis Rokas², Paula Gonçalves¹*
Variation and selection on codon usage bias across an entire subphylum

Abigail L. LaBella, Dana A. Opulente, Jacob L. Steenwyk, Chris Todd Hitinger, Antonis Rokas

Signatures of optimal codon usage in metabolic genes inform budding yeast ecology

Abigail Leavitt LaBella, Dana A. Opulente, Jacob L. Steenwyk, Chris Todd Hitinger, Antonis Rokas

The evolution of the GALactose utilization pathway in budding yeasts

Marie-Claire Harrison, Abigail L. LaBella, Chris Todd Hitinger, and Antonis Rokas

Repeated horizontal gene transfer of GALactose metabolism genes violates Dollo’s law of irreversible loss

Max A. B. Haase, Jacek Kominik, Dana A. Opulente, Xing-Xing Shen, Abigail L. LaBella, Xiaofan Zhou, Jeremy DeViglio, Amanda Beth Hullfachor, Cletus P. Kurtzman, Antonis Rokas, and Chris Todd Hitinger

Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota

Xing-Xing Shen, Jacob L. Steenwyk, Abigail L. LaBella, Dana A. Opulente, Xiaofan Zhou, Jacek Kominik, Yuanning Li, Marizeth Groenewald, Chris T. Hitinger, Antonis Rokas

A genome-scale phylogeny of the kingdom Fungi

Yuanning Li, Jacob L. Steenwyk, Ying Chang, Yan Wang, Timothy Y. James, Jason E. Stajich, Joseph W. Spatafora, Marizeth Groenewald, Casey W. Dunn, Chris Todd Hitinger, Xing-Xing Shen, and Antonis Rokas

An orthologous gene coevolution network provides insight into eukaryotic cellular and genomic structure and function

Jacob L. Steenwyk, Megan A. Phillips, Feng Yang, Swapneeta S. Date, Todd R. Graham, Judith Berman, Chris Todd Hitinger, Antonis Rokas
HGT is widespread in insects and contributes to male courtship in lepidopterans

Graphical abstract

HGT identification
218 insect genomes

Widespread HGT in insects
- Hemiptera: 249 genes
- Hymenoptera: 182 genes
- Coleoptera: 121 genes
- Lepidoptera: 613 genes
- Diptera: 197 genes

Horizontal gene transfer (HGT)
Species A B C D

HGT contributes to insect adaptation
- Development
  - Molting
- Environment
  - Coloration
- Immunity
- Reproduction
  - Male courtship

1,410 HGT-acquired genes

Authors
Yang Li, Zhiguo Liu, Chao Liu, ..., Antonis Rokas, Jianhua Huang, Xing-Xing Shen

Correspondence
antonis.rokas@vanderbilt.edu (A.R.), jhhuang@zju.edu.cn (J.H.), xingxingshen@zju.edu.cn (X.-X.S.)

In brief
A comprehensive resource of horizontal gene transfer (HGT) events in 218 insects acquired from non-metazoan sources provides insight into the adaptation of HGTs in insect genomes with the discovery of a functional role for the gene LOC105383139 in male courtship behavior in lepidopterans.
Acknowledgments

Hittinger lab – UW
Dana Opulente
Jacek Kominek
Emily Ubbelohde
& many others

C. Kurtzman – USDA
M. Groenewald – Westerdijk

Rokas lab – Vandy
Xing-Xing Shen
Xiaofan Zhou
Jacob Steenwyk
Abbe LaBella
Carla Goncalves
Kyle David
& many others

Budding Yeast Community
Moriya Ohkuma,
Rikiya Endoh, Masako Takashima,
Ri-ichiroh Manabe, Neža Čadež,
Diego Libkind, Carlos A. Rosa and many others

http://www.rokaslab.org/ @RokasLab

http://y100plus.org
Questions?

Yeasts are powerful for studying how genetic variation transforms to phenotypic variation at macroevolutionary scales

Gene conservation is not (never?) universal

Trait & gene loss are “creative” evolutionary forces

Few, but highly impactful, HGT events