Applications of genome alignment

- Comparing different genome assemblies
- Locating genome duplications and conserved segments
- Gene finding through comparative genomics
- Analyzing pathogenic bacteria against their harmless close relatives
Pipmaker

- http://www.bx.psu.edu/miller_lab/

- Visualization of BLASTZ alignments

- Although pips are compact and informative, they do not show alignment information for the second sequence.
  - Dotplots are used to see relevant differences
Maximal Unique Matches

WINDOW$  INDIGO$
1234567  1234567

Left-extensible
By “I”
Suffix links are in green

From Delcher et al., 2002, Nucleic Acids Res 30(11):2478-83
Two different most parsimonious scenarios that transform the order of the 11 synteny blocks on the mouse X chromosome into the order on the human X chromosome.

Pevzner P., Tesler G. PNAS 2003;100:7672-7677
Homology map

We multiply align these blocks together
1. Find multi-MUMs in sequences
2. Calculate pairwise distances
3. Run neighbor-joining to get guide tree;
Suffix arrays

• Suffix arrays require even less space than a suffix tree

• Very simply, it is a sorted list of suffixes
  – Example in the Aluru chapter posted as a resource
Linear time of suffix arrays

• There were three papers in 2002 that solved the old problem of constructing suffix arrays in linear time.

• These were:
  – Ko and Aluru – very interesting, but hard to understand
  – Kim et al. – was based on older parallel suffix tree algorithms
  – Karakkanen and Sanders is the simplest and most elegant.
FIGURE 1.1: Suffix tree, suffix array and Lcp array of the string mississippi. The suffix links in the tree are given by $x \rightarrow z \rightarrow y \rightarrow u \rightarrow r$, $v \rightarrow r$, and $w \rightarrow r$. 
Try it out (other way)

• Construct the suffix array of the string “BANANA$”

• Construct the LCP array for the suffix array above

• Given the suffix array and LCP array, can you draw a suffix tree?
Algorithm

- Recursively sort the \( \frac{2}{3}n \) suffixes with \( i \mod 3 \neq 0 \)

- Sort the \( \frac{1}{3}n \) suffixes with \( i \mod 3 = 0 \) using the previous result.

- Merge the two sorted arrays.
Some thoughts

• The sorting can be done using Radix sort and the relative ranks of suffixes used for the ordering.

• The 1/3 and 2/3 split makes the merging much easier; other ½ ½ approaches (e.g. Kim et al.) use this with clever tricks.

• Similar to the odd and even suffix technique of Farach.
Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118

Maité Novo, Frédéric Bigey, Emmanuelle Beyne, Virginie Galeote, Frédéric Gavory, Sandrine Mallet, Brigitte Cambon, Jean-Luc Legras, Patrick Wincker, Serge Casaregola, and Sylvie Dequin

* Author Affiliations

Edited by Jasper Rine, University of California, Berkeley, CA, and approved August 3, 2009

M.N. and F.B. contributed equally to this work. (received for review May 5, 2009)
Methods

• “Basic alignments were routinely performed with BLAST (10). Genomes were aligned, using MUMmer 3.0 (11).”

• “Gene prediction for the EC1118 genome was carried out with GlimmerHMM (12), trained on S288c ORFs from the Saccharomyces Genome Database (version 2008-10-10).”
Chromosomal distribution of the 3 unique EC1118 regions.

Region A

Region B

Region C

EC1118 unique chromosomal regions

Novo M et al. PNAS 2009;106:16333-16338
Background for yeast study

• Brewing evolved in middle ages Europe to produce ale-type beer via *Saccharomyces cerevisiae*, the same yeast used in wine and leavened bread.

• Lager-brewing arose in 15\textsuperscript{th} century Bavaria, and is the most popular technique

• Lager, however, requires slow, low temperature fermentation by cryotolerant yeast(s).
Saccharomyces pastorianus

- Used to make lager, but never has been found in wild and depends on humans.

- Allotetraploid hybrid of S. cerevisiae and an unknown yeast species.

- Understanding this unique contribution is important for understanding domestication of this yeast for human use.
Lager paper

• Three cool facts when you get a chance to read
  – Yeast used for lager beer probably arose in ale breweries
  – Two distinct types of lager yeast, referred to as groups 1 and 2
  – Both groups probably arose independently in Europe
<table>
<thead>
<tr>
<th>Strain</th>
<th>Culture collection aliases*</th>
<th>Earliest collection entry date</th>
<th>Other information</th>
<th>Collection locale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>S. pastorianus strains</strong>&lt;br&gt;Group 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSY509</td>
<td>2440</td>
<td>398</td>
<td>June 1952</td>
<td>Brewery-Saaz type beer; bottom yeast</td>
</tr>
<tr>
<td>GSY133</td>
<td>1486</td>
<td>6258&lt;sup&gt;a&lt;/sup&gt;</td>
<td>397</td>
<td>June 1935</td>
</tr>
<tr>
<td>GSY501</td>
<td>1174</td>
<td>6047&lt;sup&gt;a&lt;/sup&gt;</td>
<td>392</td>
<td>June 1931</td>
</tr>
<tr>
<td>GSY131</td>
<td>1538</td>
<td>6047&lt;sup&gt;a&lt;/sup&gt;</td>
<td>October 1935&lt;br&gt;(described by Hansen in 1904)</td>
<td>S. pastorianus-type strain</td>
</tr>
<tr>
<td>GSY137</td>
<td>6284</td>
<td></td>
<td></td>
<td>AJL248</td>
</tr>
<tr>
<td>GSY129</td>
<td>1513</td>
<td>6033&lt;sup&gt;a&lt;/sup&gt;</td>
<td>396</td>
<td>October 1947&lt;br&gt;(original culture 1883, Hansen)</td>
</tr>
<tr>
<td>GSY134</td>
<td>1503</td>
<td>6261</td>
<td></td>
<td>S. monacensis-type strain</td>
</tr>
<tr>
<td><strong>Group 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSY132</td>
<td>1260</td>
<td>6257</td>
<td>400</td>
<td>March 1937</td>
</tr>
<tr>
<td>GSY138</td>
<td>6285&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td>M 1563</td>
<td>Copenhagen</td>
</tr>
<tr>
<td>GSY139</td>
<td>6560&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td>C83 1562</td>
<td>Labatt Brewery, Canada; bottom-fermenting</td>
</tr>
<tr>
<td>GSY135</td>
<td>6282&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td>1962</td>
<td>BK 2233</td>
</tr>
<tr>
<td>GSY136</td>
<td>6283&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td>1969</td>
<td>BK 2230</td>
</tr>
<tr>
<td>GSY516</td>
<td>6903</td>
<td></td>
<td>September 1976</td>
<td></td>
</tr>
<tr>
<td>GSY515</td>
<td>5832</td>
<td></td>
<td>October 1967</td>
<td></td>
</tr>
<tr>
<td>GSY503</td>
<td>1483</td>
<td></td>
<td>July 1927</td>
<td></td>
</tr>
<tr>
<td>GSY504</td>
<td>1484</td>
<td></td>
<td>February 1925</td>
<td></td>
</tr>
<tr>
<td><strong>S. cerevisiae strains</strong>&lt;br&gt;Ale strains</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSY161</td>
<td>2156</td>
<td>457</td>
<td>June 1955</td>
<td></td>
</tr>
<tr>
<td>GSY708</td>
<td>6281&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>Wyeast1056</td>
</tr>
<tr>
<td>GSY934</td>
<td>6285&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>Leinenkugel Ale</td>
</tr>
</tbody>
</table>
Results

• Saccharomyces are associated with oak trees in Northern hemisphere.

• This study focused on Patagonia in South America with 123 cryotolerant species and two isolates of S. cerevisiae. The fact so many were cryotolerant is unique relative to the northern hemisphere.

• These group with biological assays with the two known contaminants of lager/cider/wine fermentation
Domestication and analysis

• Lager yeast is a mix of at least three yeast species

• Interestingly, all cryotolerant species have the same chunk of S. cervisiae useful for processing maltose
  – Maltose is one of the most abundant sugars in wort used in brewing

• Relationships are contentious as the lager yeast and related yeasts previously were only found in human fermentation efforts- resolved via seq