

Comparative Fungal Genomics and Core Fungal Genes

Dr. Tom Hsiang

Dept. Environmental Biology

University of Guelph

Guelph, Ontario, Canada

Presented at the 2004 Annual Meeting of the Canadian Phytopathological Society, in Ottawa, Ontario, June 13-16, 2004.

Genomic sequencing

- 1995: first prokaryote sequenced
 - *Haemophilus influenzae* (2 Mb)
- 1997: first eukaryote sequenced (fungal)
 - *Saccharomyces cerevisiae* (12 Mb)
- 1998: first animal sequenced (nematode)
 - *Caenorhabditis elegans* (97 Mb)
- 2000: first plant sequenced
 - *Arabidopsis thaliana* (115 Mb)

Genomic sequencing - plants

- 2002, *Oryza sativa* (rice, 400 Mb)

- Beijing Genomics Institute

- » var. *indica*

- International Rice Genome Sequencing Project

- » var. *japonica*

Genomic sequencing - animals

- 2000, *Drosophila melanogaster* (fly, 138 Mb)
- 2001, *Homo sapiens* (human, 2.9 Gb)
- 2002, *Fugu rubripes* (fish, 330 Mb)
 - *Ciona intestinalis* (sea squirt, 160 Mb)
 - *Mus musculus* (mouse, 2.5 Gb)
 - *Anopheles gambiae* (mosquito, 278 Mb)
- 2003, *Caenorhabditis briggsae* (worm, 104 Mb)
- Rat, cow, chimpanzee, bee almost complete

Genomic sequencing - fungi

- 2001, *Aspergillus fumigatus* (29 Mb)
- 2002, *Phanerochaete chrysosporium* (30 Mb)
- 2002, *Candida albicans* (15 Mb)
- 2002, *Magnaporthe grisea* (38 Mb)
- 2002, *Schizosaccharomyces pombe* (14 Mb, complete)
- 2003, *Aspergillus nidulans* (30 Mb)
- 2003, *Gibberella zeae* (36 Mb)
- 2003, *Cryptococcus neoformans* (39 Mb)
- 2003, *Neurospora crassa* (21 Mb, complete)

Genomic sequencing - fungi

- 2003, *Coprinus cinerea* (37 Mb)
- 2003, *Trichoderma reesei* (33 Mb)
- 2004, *Phakopsora pachyrhizi* (350 Mb)
- 2004, *Podospora anserina* (36 Mb)
- 2004, *Ashbya gossypii* (9 Mb, complete)

- other fungal genomes held by private companies

Objectives

- Find common fungal genes (homologs) by comparing each yeast gene to 13 other fungal genomes
- Compare these core fungal genes to other organisms to uncover genes found only in fungi

Method of gene comparison

- Yeast genome 12 Mb -> 6,000 predicted genes
- Use Standalone TBLASTN to find homologs
 - download program, create personal databases

yeast protein: MYYIMFLYNMLLIILIFYSI...



fungus protein: MREIVHLQTLIIILIFYS.....

translate (6-frame)

fungus genome: gttcaccttcagaccggccagtggtgtaagt.....

6356 Yeast predicted genes

3340 Yeast genes with homologs in 12 of 14 fungi

Ascomycetes

| | |
|------------------------------|------------------------------------|
| <i>Aspergillus fumigatus</i> | <i>Magnaporthe grisea</i> |
| <i>Aspergillus nidulans</i> | <i>Neurospora crassa*</i> |
| <i>Candida albicans</i> | <i>Podospora anserina</i> |
| <i>Gibberella zeae</i> | <i>Trichoderma reesei</i> |
| | <i>(Saccharomyces cerevisiae*)</i> |

Basidiomycetes

| | |
|------------------------------------|------------------------------|
| <i>Cryptococcus neoformans</i> | <i>Phakopsora pachyrhizi</i> |
| <i>Phanerochaete chrysosporium</i> | <i>Coprinus cinerea</i> |
| <i>Ustilago maydis</i> | |

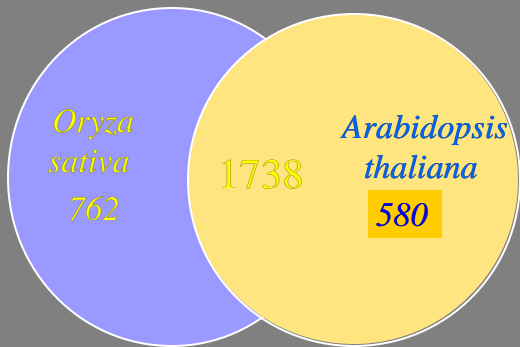
3340 genes common to 12 of 14
fungi compared to plant genomes

Oryza sativa
2500



Arabidopsis
thaliana
2318

3340 genes common to 12 of 14
fungi compared to plant genomes



Total: 3080

3340 genes common to 12 of 14
fungi compared to animal genomes

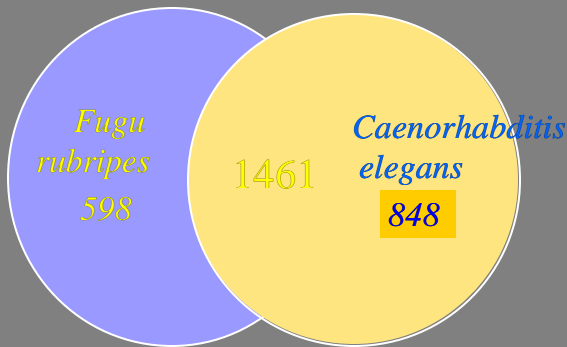
Fugu rubripes
2059



Caenorhabditis
elegans
2309



3340 genes common to 12 of 14
fungi compared to animal genomes



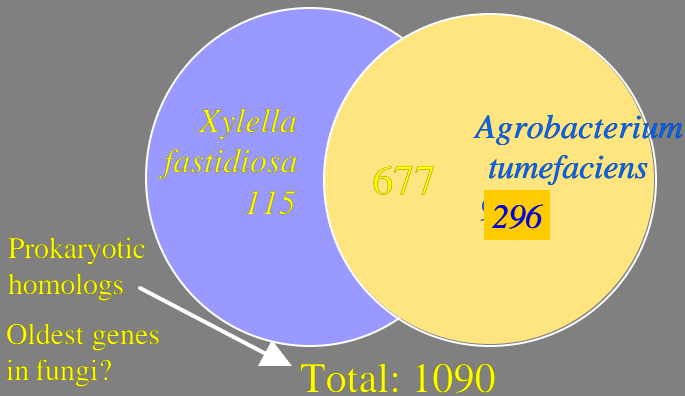
Total: 2907

3340 genes common to 12 of 14
fungi compared to bacterial genomes

Xylella fastidiosa
792

*Agrobacterium
tumefaciens*
973

3340 genes common to 12 of 14
fungi compared to bacterial genomes



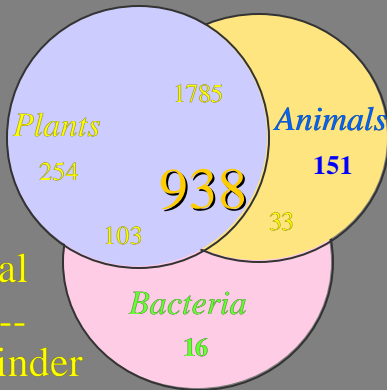
3340 genes common to 12 of 14
fungi compared to other genomes

Plants
3080

Animals
2907

Bacteria
1090

3340 genes common to 12 of 14 fungi compared to other genomes



3280 Total

60 Remainder

What are these 60 fungal genes?

- are they really found only in fungi?
 - sent to BLAST against GenBank NR , EST, GSS, and HTGS databases
- 43 have a non-fungal match at $E () \leq 10^{-5}$ so these are not exclusively fungal (many of these have homologs in the human or mouse genomes)
- this leaves 17 genes found only in fungi

What are these 17 fungal genes?

- 5 have unknown function
- 2 involved in protein biosynthesis
- 2 involved in transport
- 7 have miscellaneous functions
- 1 involved in sporulation (clearly fungal)

Note: functional annotation from Yeast Genome Database

Methods

- set up Linux Operating System (free)
 - also works in Windows, but slower
- download genomic data from internet (free)
 - see my web copy for details
- set up Standalone Blast program (free)
 - from www.ncbi.nlm.nih.gov/BLAST/
- learn PERL script programming (free but takes much time for biologists)

Uses for Comparative Genomics

- phylogenetics (whole genome comparisons)
- targeted drugs (pathogen gene absent in host)
- finding genes (conserved sequences)
- gene function (guilt by association)
- primer design
- in silico PCR
- whatever you can dream up!

Acknowledgements

- Dr. David Baillie of Simon Fraser University for helping me with PERL and allowing me access to his computing cluster
- Dr. Zamir Punja for providing benchspace and internet access

Thank you for your attention

This seminar has been posted on the web at:
www.uoguelph.ca/~thsiang/china03/genomics.pdf

email: thsiang@uoguelph.ca

Any Questions?