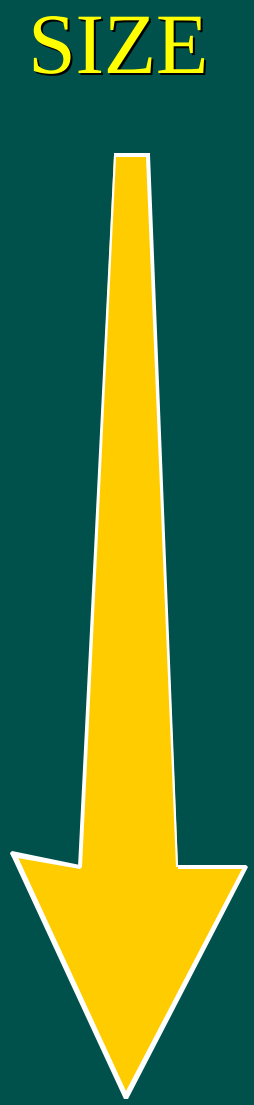
Comparative Genomics and Core Fungal Genes

Tom Hsiang
Dept. Environmental Biology
University of Guelph
Guelph, Ontario, Canada

Genomic sequencing

- 1995: first prokaryote sequenced
  - – *Haemophilus influenzae* (2 Mb)
- 1997: first eukaryote - fungus
  - – *Saccharomyces cerevisiae* (12 Mb)
- 1998: first animal - nematode
  - – *Caenorhabditis elegans* (97 Mb)
- 2000: first plant
  - – *Arabidopsis thaliana* (115 Mb)
- 2002: *Oryza sativa* (400 Mb)
- 2004: *Populus trichocarpa* (5.6 Gb)



Genomic sequencing - animals

- 2001, *Homo sapiens* (human, 2.9 Gb)
- Some insects, sea animals, mouse complete
- Rat, cow, chimpanzee, bee almost complete

Genomic sequencing - fungi

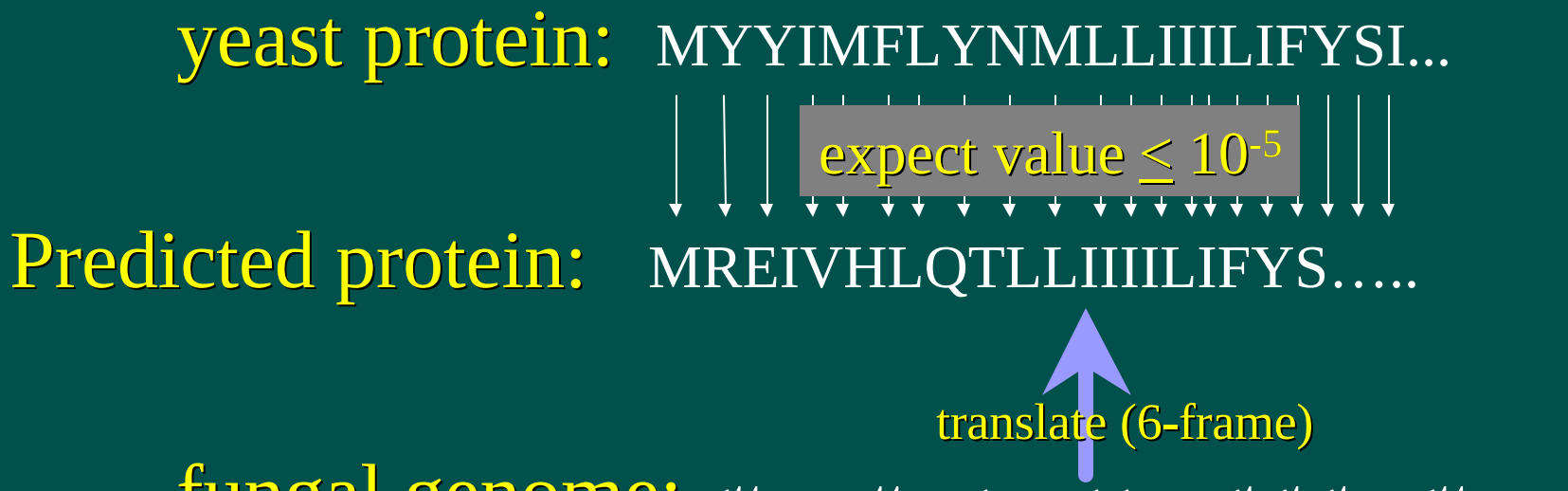
- Size of published genomes averages 30 Mb
- Over 20 species to date in publicly accessible databases
- other fungal genomes held by private companies

Objectives

- Find core fungal sequences by comparing each yeast gene to 13 other fungal genomes
- Find sequences which are unique to fungi by comparing these core fungal sequences to other organisms

Method of gene comparison

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



6356 Yeast predicted genes

3340 Yeast genes with homologs in 12 of 14 fungi

Ascomycetes

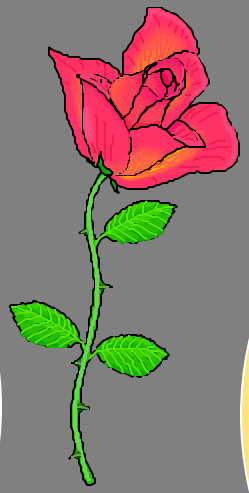
Aspergillus fumigatus
Aspergillus nidulans
Candida albicans
Gibberella zeae
Magnaporthe grisea
*Neurospora crassa**
Podospora anserina
Trichoderma reesei
(Saccharomyces cerevisiae)*

Basidiomycetes

Cryptococcus neoformans
Phanerochaete chrysosporium
Ustilago maydis
Phakopsora pachyrhizi
Coprinus cinerea

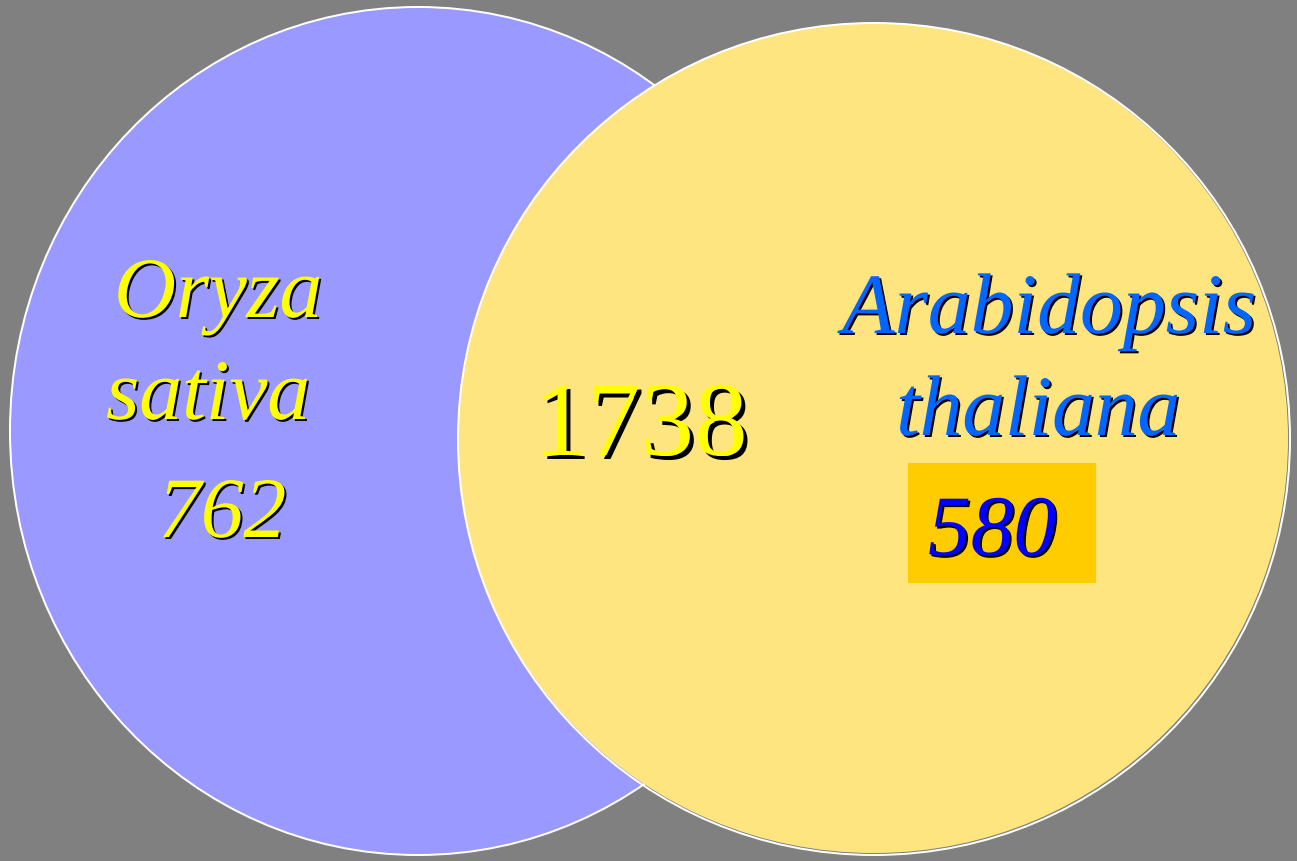
3340 genes common to 12 of 14 fungi compared to plant genomes

Oryza sativa
2500



Arabidopsis thaliana
2318

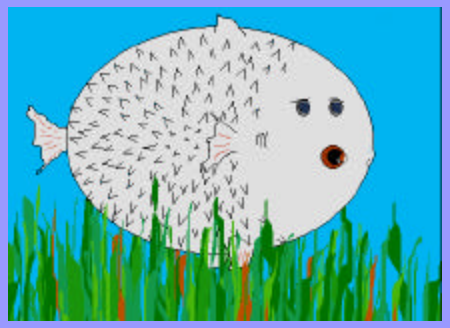
3340 genes common to 12 of 14 fungi compared to plant genomes



Total: 3080

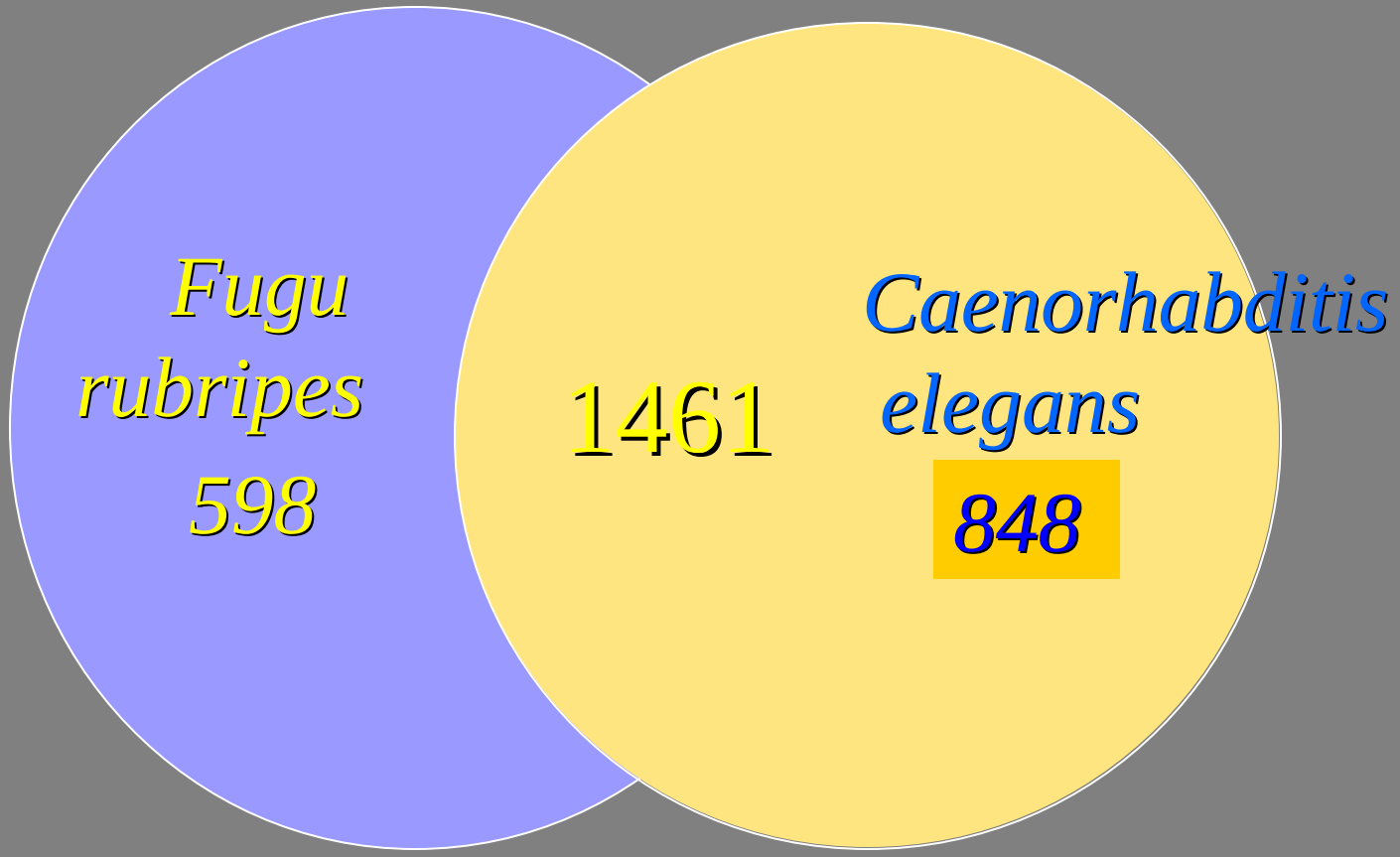
3340 genes common to 12 of 14 fungi compared to animal genomes

Fugu rubripes
2059



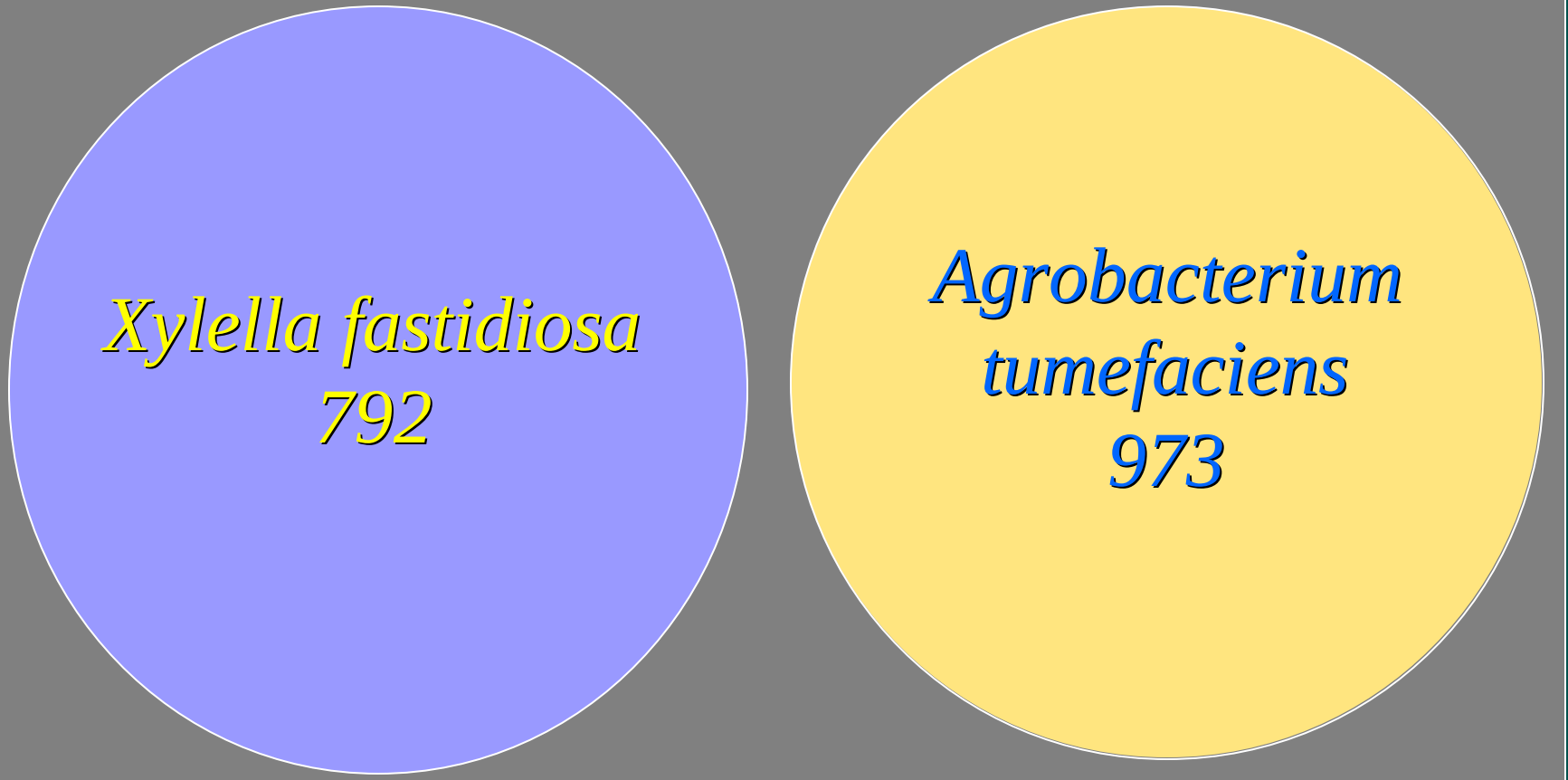
Caenorhabditis elegans
2309

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

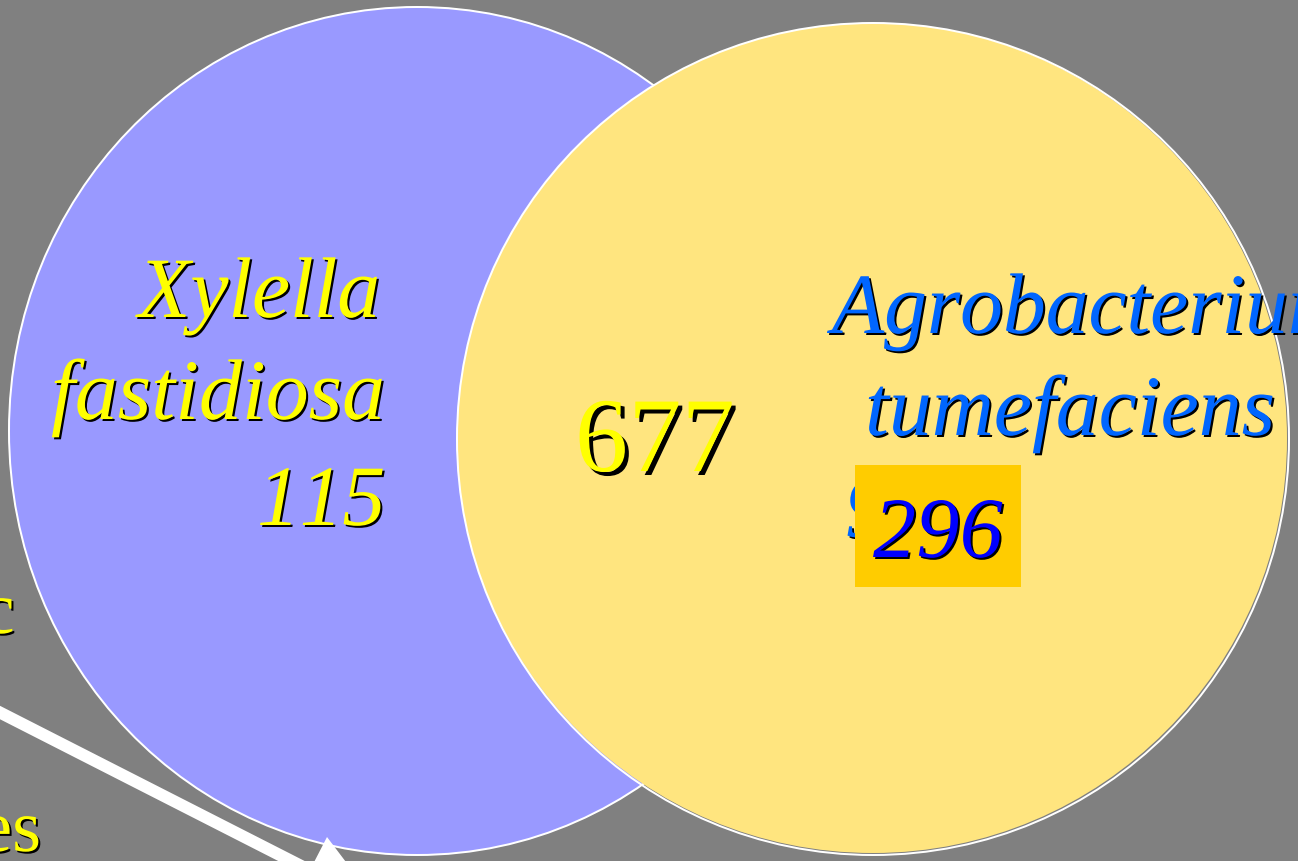


Total: 2907

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



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

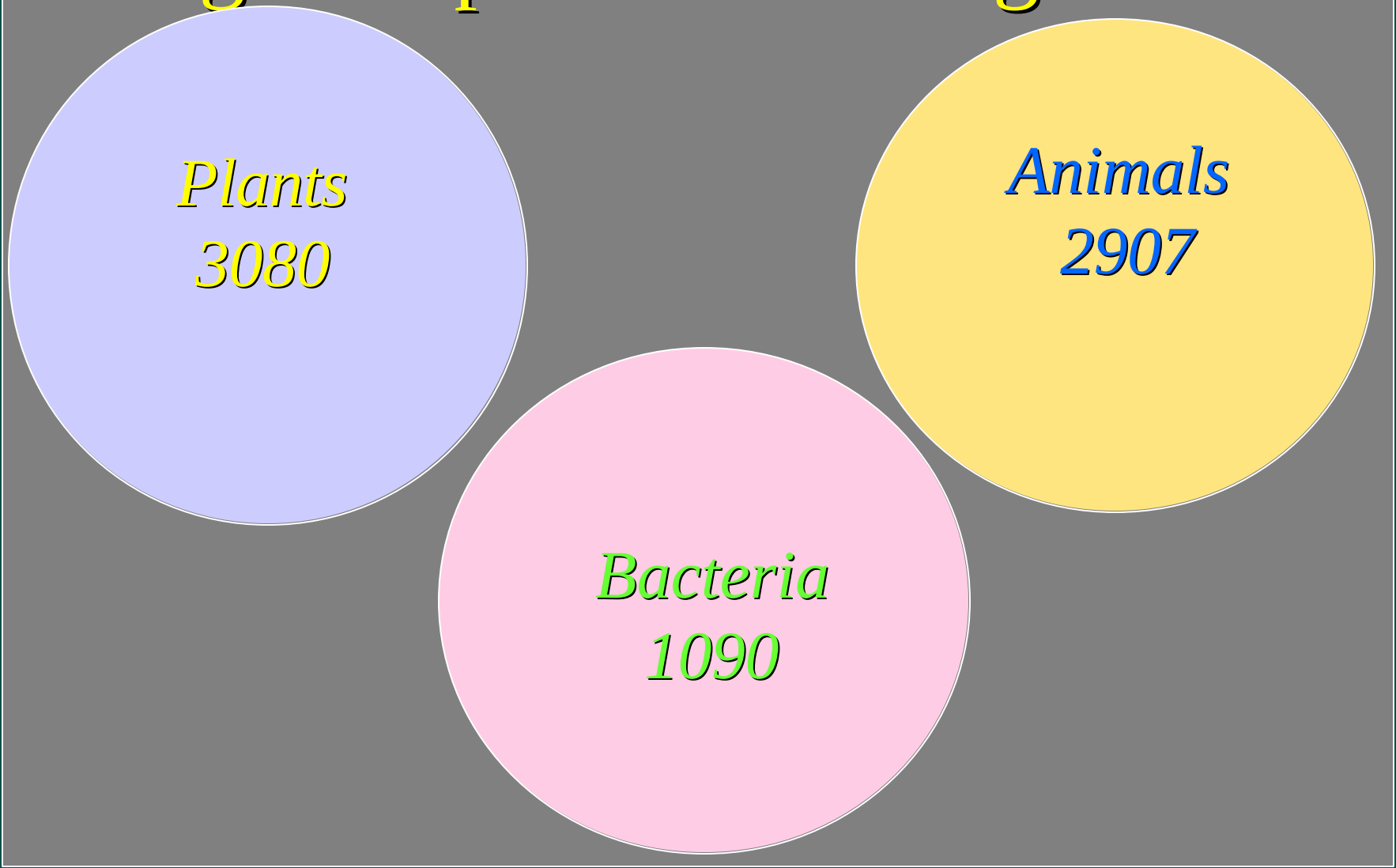


Prokaryotic homologs

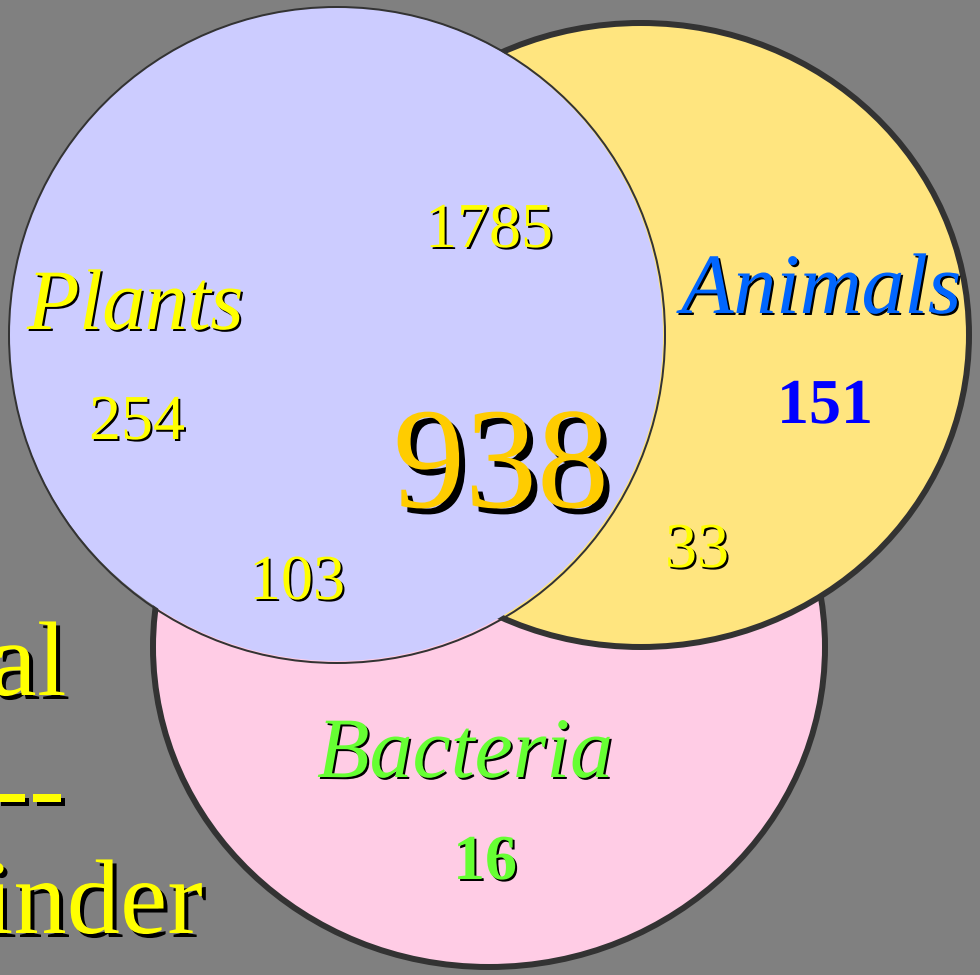
Oldest genes in fungi?

Total: 1090

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



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)