

Comparative genomics to assess horizontal gene transfer between mycorrhizal fungi and plants

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

Versions of this presentation given at the following meetings:
Plant Canada, June 2007, Saskatoon, Saskatchewan, Canada
Mycorrhizal Conference, Aug 2007, Chuxiong, Yunnan, China

Mycorrhizal associations

- Ancient?
 - Arising with the establishment of land plants?
- Intimate?
 - exchange nutrients
 - exchange genes?

Horizontal Gene Transfer (HGT)

- Gene acquisition not by descent
- Frequent among prokaryotes
 - » (transformation, transduction, conjugation)
- Thought to be rare with eukaryotes (<10 cases)
 - chloroplasts from cyanobacteria?
 - mitochondria from proteobacteria?
 - bacterial genes into host eukaryotes
 - parasitic plant gene into host plant
- Sequences from complete genomes allows us to look more deeply in the possibility of HGT

Comparative genomics

- Comparison of genome sequence data
 - Completely sequenced genomes (>95%)
 - > 600 prokaryotic genomes (Archaea & bacteria)
 - > 50 fungal genomes (especially yeasts)
 - > 20 animal genomes (insects, fishes, mammals)
 - > 10 plant genomes (arabidopsis, rice, poplar, corn, green alga, coffee)
- data from Genomes Online (<http://www.genomesonline.org/>) and other sources

Objectives

- Compare the genome of the mycorrhizal agaric, *Laccaria bicolor*, with available plant genomes and fungal genomes
 - search for homologs shared between *Laccaria* and plants (poplar, rice, arabidopsis) but not other fungi tested
- Compare the genome of a saprophytic, non-mycorrhizal agaric, *Coprinus cinereus*, with available plant genomes
 - contrast these results with the *Laccaria* results

Methods: sequence comparisons

- *Laccaria* genome 64 Mb -> 20K+ predicted genes
 - Download genomes & create local databases
 - Use Standalone BLAST to find homologs
 - BLASTP (protein vs. protein)
- Laccaria protein: MYIIMFLYNMLLIILIFYSI...
(query sequence)
- target protein: MREIVHLQTLIIILIFYS....
(subject sequence)
- Expect value $\leq 10^{-10}$
- Sequence similarity implies descent from common ancestor?
- The Expect Value refers to "the number of hits one can expect to see just by chance when searching a database of a particular size" (probability of match by chance)

```
BLAST 2.2.6 (Apr-09-2003)
Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1990),
"Basic Local Alignment Search Tool: a new generation of protein database search
programs". Nucleic Acids Res. 18:3298-3402.

Query: 12661261 (1845 letters)
Database: Database of GenBank+EMBL+DBJ sequences from EBI Databases
2,838,841 sequences/ 856,658,774 total letters

Sequences producing significant alignments:

gi|246879110|gb|EF121212.1|Mus musculus strain EP155 A... 104 46.020
gi|246879110|gb|EF121212.1|Mus musculus strain EP155 A... 85 34.000
gi|246879110|gb|EF121212.1|Mus musculus strain EP155 A... 85 34.000
gi|246879110|gb|EF121212.1|Mus musculus strain EP155 A... 85 34.000
gi|246879110|gb|EF121212.1|Mus musculus strain EP155 A... 85 34.000
gi|246879110|gb|EF121212.1|Mus musculus strain EP155 A... 85 34.000
gi|246879110|gb|EF121212.1|Mus musculus strain EP155 A... 85 34.000
gi|246879110|gb|EF121212.1|Mus musculus strain EP155 A... 85 34.000
gi|246879110|gb|EF121212.1|Mus musculus strain EP155 A... 85 34.000
gi|246879110|gb|EF121212.1|Mus musculus strain EP155 A... 85 34.000

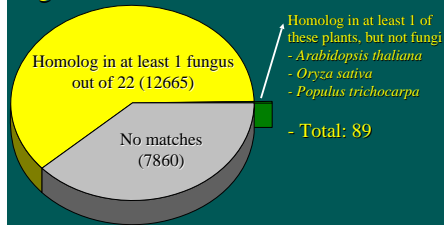
>gi|246879110|gb|EF121212.1|Mus musculus strain EP155 A...
MREIVHLQTLIIILIFYS....
*****
MREIVHLQTLIIILIFYS....
Length = 477

Expect = 1.22E-126
Positives = 43/223 (19%), Gaps = 1/223 (5%)
Query: 244 KQIYKSPHIIILILPILITLITVPIVLIYKVFVPEVETNENNESVVAKZKRRKAPY 303
          2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Sbjct: 6  KQIYKSPHIIILILPILITLITVPIVLIYKVFVPEVETNENNESVVAKZKRRKAPY 303
```

20614 *Laccaria* predicted genes

- 12665 with homologs in at least one of 22 fungi
- | | |
|-----------------------------------|------------------------------------|
| Ascomycota | Ascomycota (con't) |
| <i>Ashbya gossypii</i> | <i>Sclerotinia sclerotiorum</i> |
| <i>Aspergillus fumigatus</i> | <i>Stagonospora nodorum</i> |
| <i>Aspergillus nidulans</i> | <i>Trichoderma reesei</i> |
| <i>Botrytis cinerea</i> | <i>Ustilagocarpus reesii</i> |
| <i>Candida albicans</i> | |
| <i>Chaetomium globosum</i> | Basidiomycota |
| <i>Gibberella zeae</i> | <i>Phanerochaete chrysosporium</i> |
| <i>Magnaporthe grisea</i> | <i>Coprinus cinereus</i> |
| <i>Nectria haematococca</i> | <i>Cryptococcus neoformans</i> |
| <i>Neurospora crassa</i> | <i>Ustilago maydis</i> |
| <i>Podospora anserina</i> | Zygomycota |
| <i>Saccharomyces cerevisiae</i> | <i>Rhizopus oryzae</i> |
| <i>Schistosaccharomyces pombe</i> | |

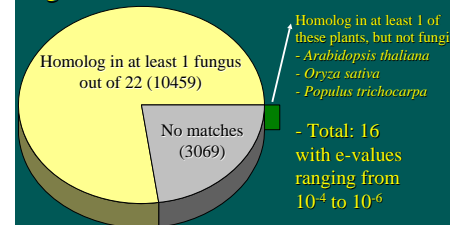
20614 *Laccaria bicolor* predicted genes



89 *Laccaria* predicted genes with homologs in plants but not fungi compared to GenBank NR: top hits

- no matches: 17
- other fungi: 16
- animals: 16
 - e-value < 10⁻¹⁰: 12
- bacteria: 8
 - e-value < 10⁻³⁰: 7
- protists: 4
 - e-value < 10⁻³⁰: 4
- other: 2
- plants et al.: 26
- plants only: 19
 - e-value < 10⁻¹⁰: 12
 - e-value < 10⁻³⁰: 7

13544 *Coprinus cinereus* predicted genes



Summary

- 77% of *Coprinus cinereus* predicted genes vs. 61% of *Laccaria bicolor* predicted genes had a fungal match in Standalone BLAST against 22 fungal & 3 plant genomes
- 16 predicted genes of *Coprinus cinereus* had only a plant match in a test against 22 fungi and 3 plants genomes, vs. 89 predicted genes of *Laccaria bicolor*
- perhaps some of these resulted from HGT
- ongoing lab work to verify these findings