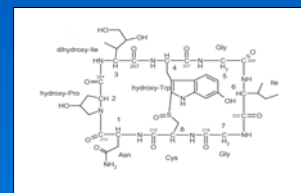




Examining amatoxins: The *Amanita* Genome Project

Heather E. Hallen* & Jonathan D. Walton.

MSU DOE Plant Research Laboratory, Michigan State University, East Lansing, MI 48824-1312, USA. *hallenhe@msu.edu.



Abstract

As a pilot study for the whole genome, we have sequenced 9,560 random genomic clones from *Amanita bisporigera*, a total of 5.7 Mb. This information is publicly available at <<http://www.pr.lmsu.edu/walton/amanita.htm>>.

Introduction

Basidiomycete fungi are relatively unrepresented in genome projects. *Amanita* is important because of its obligate mycorrhizal nature. The ectomycorrhizal basidiomycete *Laccaria bicolor* was approved for genome sequencing in 2004 with completion expected this year, and several EST projects are underway, but the ectomycorrhizal fungi are still very far behind pathogenic or industrial fungi in terms of sequencing efforts. An additional feature that makes *Amanita* of interest is the suite of secondary metabolites produced by the genus. Genome sequence would facilitate studies on the ecology, physiology, biochemistry and evolution of this important group of fungi.

Our particular interest is in identifying the genes involved in amatoxin biosynthesis. Amatoxins are presumed to be synthesized by a non-ribosomal peptide synthetase (NRPS) (Walton *et al.* 2004). With eight amino acids, amatoxins should require an eight-module NRPS of approximately one MDa. Such a protein is predicted to be encoded by a 30-kb gene. Assuming random sampling across the genome and an average read size of 600 bp, there is > 99% probability of hitting a 30 kb target in a 40 Mb genome in 7,000 random, independent sequences.

We chose shotgun genome sequencing after extensive unsuccessful attempts to obtain amatoxin synthetase through PCR and biochemical methods (ATP-pyrophosphate exchange assay; amino acid feeding studies). A cDNA sequencing project was considered impracticable, as there is strong evidence suggesting that amatoxins are synthesized only during a small window of time, during or immediately preceding button initiation (Preston *et al.* 1982; Hallen pers. obs.). We have now generated 7414 contigs, assembled from 9560 genomic sequence reads, from *Amanita bisporigera*. Each sequence has been compared to GenBank's non-redundant database as well as the *Coprinopsis* and *Phanerochaete* genomes. We maintain all sequences in a BLAST-searchable format.

Methods

Amanita bisporigera was selected for this study due to the relative ease in obtaining specimens and the consistent high levels of amatoxins and phallotoxins found in *A. bisporigera* fruiting bodies. Two DNA libraries consisting of 2-kb fragments were generated, one by the Genomics Technology Support Facility at Michigan State University and one by MacroGen, Inc. Several thousand clones from each library were sequenced and BLAST searches were performed. BLASTX (translated query against protein database) was used in searching the non-redundant database (NR) at GenBank, and TBLASTX (translated query against translated database) and BLASTN (nucleotide query against nucleotide database) were used in searching the genomes of *Coprinopsis* and *Phanerochaete*, the two closest relatives to *Amanita* for which complete genome sequence was available.

BLAST results were examined, catalogued, and automatically annotated. We analyzed only those hits with expect values of greater than 1.00E-39. Hits with higher (=worse) expect values were considered tenuous and potentially misleading.

Results and Discussion

A breakdown by functional category of reliable hits to NR is shown in Fig. 1.

The largest classes of genes are those devoted to metabolism (30.3%) and cellular processes (24.7%). These percentages are similar to those observed in *Neurospora crassa* in which 23% of annotated genes pertain to metabolism and 27% to cellular processes (Schulte 2004).

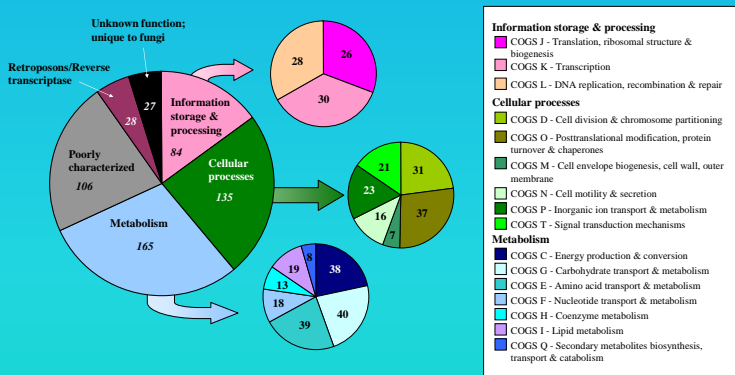


Fig. 1. A breakdown of genes identified by BLASTX hits with expect values equal to or less than 1.00e-39. Genes are functionally categorized using COGS (Clusters of Orthologous Groups) as implemented by NCB (Tatusov *et al.* 2003). The large pie chart on the left shows the supercategories, which are further broken down in the three smaller pie charts to the right. Numbers correspond to the number of *Amanita* sequences identified as belonging in each category. Colors in the smaller pie charts correspond to the functional categories in the legend on the right.

Query	Hit Description	Bit Score	E-value
16_c01KoreaM13Rrc	alpha-aminoadipate reductase [Cryptococcus neoformans var. neoformans]	349	5.0E-95
W7_25_H11	symbiosis-related protein [Laccaria bicolor]	101	1.0E-31
cn732	Ich1 [Coprinopsis cinerea] [gi7493995]pir1[T00249] ich1 protein - inky cap (Copr	176	3.0E-57
W7_39_A02	Cap64 protein [Pleurotus ostreatus]	306	8.0E-82
W7_85_G10	laccase [Auricularia polytricha]	127	3.0E-46
w7_13_xp_G04	laccase 2 precursor [Coprinus cinereus] [gi37703767] [gb AA01243.1] laccase	116	1.0E-45

Fig. 2. A small sampling of genes of interest that have been detected in the *Amanita bisporigera* genome.

Fig. 2 shows a sample of preliminary results of the BLASTX searches against GenBank's non-redundant protein database.

- L-aminoadipate-semialdehyde dehydrogenase is related to non-ribosomal peptide synthetases, the enzymes believed to be responsible for amatoxin synthesis. We sequenced the remainder of the clone 16_c01KoreaM13Rrc and extended the sequence by approximately 700 bp using inverse PCR before we were able to rule out the possibility of its being a non-ribosomal peptide synthetase.
- Very little is known as yet about the genetic foundation of the ectomycorrhizal symbiosis. The hit to the symbiosis-related protein from another ectomycorrhizal fungus, *Laccaria bicolor*, is therefore of interest. We expect the *Amanita* genome and the *Laccaria* genome (expected completion 2005) to provide mutual support in identifying the genes responsible for initiating and maintaining the symbiosis.
- Ich1 is essential for pileus formation in *Coprinopsis cinerea* (Muraguchi & Kamada, 1998).
- Cap64 is a capsule formation protein first identified in the pathogenic basidiomycete *Filobasidiella neoformans*. It possesses a homolog in the saprophytic basidiomycete *Pleurotus ostreatus*. The function of Cap proteins in *Amanita* and *Pleurotus*, which do not form the capsules associated with mammalian pathogenicity, is unknown.
- Laccases are widespread in saprophytic fungi (*Coprinopsis*, *Melanocarpus*, and the white rot fungus *Trametes*), both ascocand basidiomycetes. Their role in an ectomycorrhizal fungus, which is presumably obtaining most of its nutrients in the form of photosynthate and should therefore lack the need to degrade plant tissue, is unknown.

Approximately 59% of our *Amanita bisporigera* sequences have no hit to the GenBank NR database. This is consistent with other fungal genome projects (see, e.g. Schulte 2004); little annotation is yet available for fungal genomes, so the proportion of unidentified sequences is high. 3008 sequences produce no hits to NR, but do yield hits to the *Phanerochaete* and/or *Coprinopsis* genomes, which are not in NR.

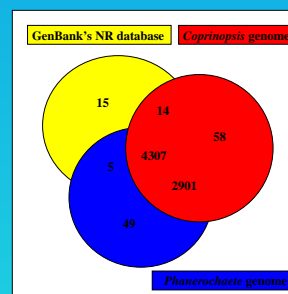


Fig. 3. Distribution of BLAST hits to NR (yellow), *Coprinopsis cinereus* (red) and *Phanerochaete chrysosporium* (blue) (not to scale).

There were 4341 hits to GenBank's NR database, 7262 to the *Phanerochaete chrysosporium* genome and 7280 to the *Coprinopsis cinereus* genome. An additional 65 *Amanita* sequences did not yield any hits to NR, *Coprinopsis* or *Phanerochaete*. 4307 hits were common to NR, *Coprinopsis* and *Phanerochaete*, while another 2901 sequences produced hits in *Coprinopsis* and *Phanerochaete* but not in NR (Fig. 3).

If we refine the analysis of *Coprinopsis* and *Phanerochaete* by limiting it to hits with an expect score at or below 1.00E-4, we are left with 3609 hits to *Phanerochaete* and 3845 hits to *Coprinopsis*.

The *Amanita* Genome and You

We would like our genome to be widely available and widely used by the scientific community. We consider our current work a prelude to a full-scale genome sequencing project. We already possess approximately 15% of the genome in a searchable and accessible format. The existing data have many potential applications. Our sequences have been used to design microsatellite primers for a study in the related *Amanita phalloides*.

Researchers studying particular genes are already using the BLAST-searchable *A. bisporigera* database to look for homologs to their genes. To date, the *Amanita* genome is best accessed through the *Amanita* Genome Project webpage, <<http://www.pr.lmsu.edu/walton/amanita.htm>>. We invite you to access the webpage, and we welcome any suggestions or comments.

References

- Muraguchi H, T Kamada (1998) The ich1 gene in the mushroom *Coprinus cinereus* is essential for pileus formation in fruiting. *Development* 125: 3133-3141
- Preston JF, BEC Johnson, M Little, T Romeo, HJ Stark, JE Mullersman (1982) Investigations on the function of amatoxins in *Amanita* species: a case for amatoxins as potential regulators of transcription. In *Peptide Antibiotics - Biosynthesis and Functions* (H Kleinkauf & H von Döhren, eds.), 399-426. Walter de Gruyter, Berlin.
- Schulte, U (2004) Genomics of filamentous fungi. In *Advances in Fungal Biotechnology for Industry, Agriculture, and Medicine* (S Tkacz & L Lange, eds.): 15-29. Kluwer Academic/Plenum Publishers, New York.
- Tatusov RL, ND Fedorova, JD Jackson, AR Jacobs, B Kiryutin, EV Koonin, DM Krylov, R Mazumder, SL Mekhedov, AN Nikolskaya, BS Rao, S Smirnov, AV Sverdlov, S Vasudevan, YI Wolf, JI Yin, DA Natale (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4: 41
- Walton JD, DG Panaccione, HE Hallen (2004) Peptide synthesis without ribosomes. In *Advances in Fungal Biotechnology for Industry, Agriculture, and Medicine* (S Tkacz & L Lange, eds.): 127-162. Kluwer Academic/Plenum Publishers, New York.