IIp-1

A renaissance in fungal genetics: progress and research at the Fungal Genetics Stock Center

Kevin McCluskey*, Sheera Walker, Rachel Yedlin, Mike Plamann

Fungal Genetics Stock Center, School of Biological Sciences, University of Missouri- Kansas City

Being at the FGSC gives us opportunities to conduct studies that rely heavily on the resources in the collection. One such project is an attempt to associate open reading frames with otherwise anonymous genetic markers in *Neurospora*. Since temperature sensitive alleles are unlikely to result from the deletion mutagenesis project being carried out by the *Neurospora* Functional Genomics Program Project we determined to identify ORFs of TS strains in *Neurospora*. Nearly 50 such markers exist in *Neurpospora crassa* and they are known as "unknown" strains.
We have complemented two such markers using cosmid walking from linked markers.

In 2005 the FGSC distributed nearly two times as many strains as ever before. We sent out 2125 fungal strains to 138 different recipients in thirty-one different countries. Not including KO strains, we added 406 new strains to the collection. 277 of these were *Neurospora* strains. Six were other organisms that have been sequenced or will be soon. We have also added a number of new resources to the FGSC web-site including all back issues of the Aspergillus newsletter. We also built an interface that allows us to link between online methods and existing indices. Overall, the resources at the FGSC are being used by investigators all over the world for a variety of purposes.

The MIPS comprehensive fungal genome resources

Ulrich Güldener[1*], Gertrud Mannhaupt[2], Martin Münsterkötter[1], Hans-Werner Mewes[1]

[1] National Research Centre for Environment & Health, 85764 Neuherberg, Germany, +49 89/3187-3579, u.gueldener@gsf.de
[2] Department of Organismic Interactions, Max Planck Institute for terrestrial Microbiology, Karl-von-Frisch Straße, D-35043 Marburg, Germany

The MIPS fungal resources focus on manually curated genome data sets as well as on genomes analyzed with automated systems only. Beside the CYGD (Comprehensive Yeast Genome Database) we set up and maintain manually annotated databases on three filamentous fungi: *Neurospora crassa* (MNCDB), *Fusarium graminearum* (FGDB) and the basidiomycete *Ustilago maydis* (MUMDB). For analysis of the protein data we use the PEDANT system as well as the newly developed SIMAP resource; both provide similarity data as well as protein domain information and structural analysis and include the analysis of additional 36 yeasts and further filamentous species. The InterPro data is used for a first inter-species comparative view.The data are organized in the well established MIPS catalogs (e.g. FunCat) and novel query techniques are available to search any field in the databases. The query interface is similar to the Entrez service at NCBI and will be further developed to enable comparative queries across species. As a special repository for protein interaction data, the MIPS Protein interaction database (MPact) was developed, which currently holds the yeast interaction data, also used for mapping of this data to other species (interologs). The created resources establish the basis for comparative research to explore the pathogenic factors in the plethora of plant and animal fungal pathogens, as well as targets of biotechnological interest.

IIp-3

Microarrays and gene silencing for functional genomics of *Sclerotinia sclerotiorum*

Adrienne Sexton[1], David Edwards[2], Megan Drew[3], Donald Gardiner[1], Barbara J Howlett[1], Hayley Ridgeway[4], Richard Weld[4], Kim M Plummer[3*], Alicia Greenhill[1]

[1] *School of Botany, University of Melbourne*
[2] *Primary Industries Research Victoria, Department of Primary Industries Victorian AgriBiosciences Centre Australia*
[3] *Botany Department, La Trobe University*
[4] *National Centre for Advanced Bio-Protection Technologies Lincoln, New Zealand*

Sclerotinia sclerotiorum is a devastating pathogen that causes stem rot of many plants, including major crop species. We are utilising the genome database for *S. sclerotiorum*. (Broad Institute) to identify genes likely to be involved in infection and survival of this fungus. Combimatrix oligonucleotide microarrays, with 12,000 probes, are being designed to analyse *S. sclerotiorum* gene expression during early stages of infection of *Brassica napus* (canola). Quantitative RT-PCR will be used to verify expression for genes of interest. Our research focus is on secreted proteins (effector candidates) and enzymes involved in biosynthetic pathways that are up-regulated specifically during host penetration/infection. Selected genes will then be targeted for genetic analysis using gene silencing and gene disruption.

Gene silencing has been demonstrated to be an efficient tool for gene analysis in a related fungus, *B. cinerea*, however it has not been reported for *S. sclerotiorum*. Both fungi often have multiple nuclei and multiple copies of genes which can make gene disruption (by homologous integration) difficult. Gene silencing (RNA interference) overcomes this problem by targeting transcripts rather than the genomic copy of the gene itself. A biosynthetic gene for the pigment, melanin, is being targeted. *S. sclerotiorum* produces a vegetative survival structure (a sclerotium) that is protected against degradation in the soil by a tough, pigmented (melanised) coat. The sclerotia are an important source of inoculum and they survive in soil for many years. We are producing a construct, with a hairpin targeted to the THN (tetrahydroxynaphthalene) gene in a Gateway vector. Our overall aim is to establish whether gene silencing has utility for high-throughput genetic analysis of *S. sclerotiorum* genes.

Comparative genomics of fungal genomes for analysis and validation of predicted genes from high-throughput genome annotation

Igor Grigoriev*, Andrea Aerts, Alan Kuo and Asaf Salamov

US DOE Joint Genome Institute, Walnut Creek, CA, USA

We annotated genomes of *Phanerochaete chrysosporium*, *Tricoderma reesei*, *Laccaria bicolor*, *Nectria haematococca*, and several other fungi that had been recently sequenced at the US DOE Joint Genome Institute. JGI Annotation pipeline that we developed and used for annotation of these genomes integrates gene prediction, annotation, and analysis methods and is targeted to annotation of a diverse set of genomes in high-throughput but genome-specific manner. Quality of predicted gene models in each genome was evaluated on basis of homology to proteins from other organisms and available experimental evidence such as ESTs/cDNAs and proteomics. This traditional genome-centric approach allowed us to validate most of predicted genes and eliminate potential over-predicted gene models (mostly ab initio models with no additional evidence). In addition, we employed a comparative genomics approach to identify a 'core' set of functions and pathways conserved among fungi and find missed genes in each of the annotated genomes. We compared domain composition, analyzed clusters of orthologs and metabolic pathways. The results demonstrate that comparative genomic analysis made more efficient both annotation QC to find missed genes and genome analysis to discover genome-specifics gene sets.

IIp-5

The genome sequence of the dothideomycete *Stagonospora nodorum*

Richard P Oliver[1], Peter S Solomon[1*], Chinnappa Kodira[2], Bruce Birren[2]

[1] *Australian Centre for Necrotrophic Fungal Pathogens, SABC, DHS, Murdoch University, WA 6150, Australia.*
[2] *The Broad Institute, Cambridge, MA 02141-2023, USA.*

Stagonospora (syn. *Septoria*) *nodorum* is a heterothallic ascomycete and a major foliar pathogen of wheat and other cereals (Wiese, 1987). In collaboration with the Broad Institute in Boston, the genome sequence of strain SN15 has been sequenced, assembled and auto-annotated (http://www.broad.mit.edu/annotation/fungi/stagonospora_nodorum). The sequence was acquired using a random shotgun cloning approach. Approximately 320,000 plasmid end reads and 50000 fosmid end reads were obtained. Some 380,000,000 bp of sequence data was obtained representing 10× coverage. Assembly of the genome gave 496 contigs averaging 74.7 Kb (range 2.0 - 562.0 Kb). Contigs were linked via paired fosmid reads into 109 supercontigs (scaffolds) with an average length of 341.6 Kb (range 2.0 Kb - 2.5 Mb). The assembled sequenced totalled 37.1Mbp, very similar to other published filamentous fungal genome sequences. Analysis of the genome sequence thus far has uncovered many interesting features including strong evidence of RIP. These results and other notable features uncovered thus far will be presented.

This is the first publicly available genome sequence representing the Dothideomycete, an taxon including some of the most important pathogens of cereals, canola and legumes worldwide (eg. Pyrenophora, Cochliobolus, Leptosphaeria, Alternaria and Ascochyta spp.). Further analysis of the genome is hoped to shed light on the potent mechanisms of pathogenicity that these fungi possess.

Wiese, M.V. (1987) Compendium Of Wheat Diseases. St. Paul.: American Phytopathological Society.

The central *Aspergillus* data repository (CADRE): viewing and analyzing sequences

Jane Mabey Gilsenan*, Peter F Giles, Michael J Anderson, Teresa K Attwood, Stephen G Oliver, Norman W Paton, Geoffrey D Robson, David W Denning

The University of Manchester, Manchester, M13 9PT, U.K.

CADRE (http://www.cadre.man.ac.uk) has been established to house publicly available genomic data for all Aspergillus species. Using a Web browser, the user can view annotation and utilise sequence data from assembled contigs. Information that can be displayed includes the position on contigs of mapped features, such as protein-coding genes and RNA-coding genes. For each feature, links are provided to allow the user to retrieve further data. For protein-coding genes, such data include: chromosomal co-ordinates; a description of the encoded protein's function; similarity matches; and displays of transcript and protein structures. Using their own sequences, the user can also initiate a BLAST search to find matches within one or several genome(s). A summary of all matches is displayed on completion of the search. For each match, contig and chromosomal coordinates are provided, along with an ideogram indicating its chromosomal location. Hyperlinks are also displayed that enable the user to view the actual sequence alignment and the position of the match relative to a contig. Two finished Aspergillus genomes are currently available on the Website. The A. nidulans genomic sequence and annotation has been provided by the Broad Institute and consists of 28.6Mb of DNA with 9520 predicted protein-coding genes. The sequence consists of 248 contigs assembled into 8 linkage groups. The A. fumigatus sequence was determined by The Institute for Genomic Research (TIGR) and the Sanger Institute. Annotation was generated automatically and reviewed manually by TIGR. The genome consists of 28.8Mb of DNA with 9926 predicted protein-coding genes and has been assembled into 8 chromosomes. The annotation of both genomes is currently being updated and will be displayed within CADRE when made publicly available. CADRE will also display genomic sequence and annotation for another seven genomes: A. oryzae, A. clavatus, A. flavus, A. terreus, A. niger and Neosartorya fischeri, as well as another strain of A. fumigatus. Using a comparative pipeline, syntenic blocks will be identified between species pairs and will be displayed within CADRE at the chromosomal and contig levels.

IIp-7

Genome comparison of *A. fumigatus*, *N. fischeri* and *A. clavatus*

Rama Maiti[1*], Natalie Fedorova[1], Vinita Joardar[1], Jonathan Crabtree[1], Samuel V. Angiuoli[1], Michael J. Anderson[2], Paolo Amedeo[1], David Denning[2], Brian J. Haas[1], Jennifer R. Wortman[1], Owen R. White[1], William C. Nierman[1]

[1] The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20852, USA, Phone; 301-838-3559, Fax: 301-8383-0209, Email: jwortman@tigr.org
[2] University of Manchester, Manchester, UK

The genome of *Aspergillus fumigatus*, a life-threatening human pathogen, represents an important fungus of the ubiquitous but diverse Aspergillus group spanning over 200 million years of evolution. The publication of the *A. fumigatus* Af293 genome sequence along with the phylogenetically distant genomes of *A. oryzae* and *A. nidulans* was an important milestone, providing insights into aspects of cellular physiology, genome evolution and gene regulation in these fungi. In order to further extend our knowledge of *A. fumigatus,* NIAID has funded the sequencing and annotation of three more closely related genomes: *N. fischeri (A. fischerianus)*, *A. clavatus* and *A. terreus*.

Comparative genomic analysis at the level of these more closely related Aspergilli should enable improved annotation of *A. fumigatus* and may shed light on the evolution of pathogenicity in this species. We have performed preliminary analysis of genome structure which supports the previously determined phylogeny of these species, with *N. fischeri* the most related to *A. fumigatus*, with the longest, uninterrupted syntenic blocks; *A. clavatus* intermediate, with numerous local rearrangements and inversions; and *A. terreus* the most distant, with inter-chromosomal breaks common. *A. terreus* and *N. fischeri* are capable of causing disease in human patients, while *A. clavatus* cannot survive at body temperature, and is therefore not a viable pathogen. *A. clavatus* is a frequent causative agent in mycotoxin contamination of cattle feed.

Based on our improved annotation of two strains of *A. fumigatus* and comparative analysis between these genomes, we will present the predicted gene complements of *A. clavatus* and *N. fischeri*. Our preliminary annotation data shows differences in gene content. We will examine the species-specific genes of each of these genomes, polymorphisms and differential gene regulation. The comparative analysis of these species is expected to help understanding of the mechanisms of virulence, environmental adaptation and resistance to antifungal treatments.

Comparison of protein coding gene content of yeasts and filamentous fungi

Mikko Arvas[1*], Teemu Kivioja[1], Alex Mitchell[2], Markku Saloheimo[1], Steve Oliver[3], Merja Penttilä[1]

[1] VTT Biotechnology, Tietotie 2, Espoo P.O. box 1500, 02044 VTT, Finland
[2] EMBL Outstation – Hinxton, European Bioinformatics Institute, Welcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK
[3] University of Manchester, Michael Smith Building, Oxford Road, Manchester, M13 9PT, UK

The recent sequencing efforts of a number of fungal genomes have revealed that they contain many genes with little or no homology to known genes (20 % - 40 % of CDS with no hits in Interpro, www.ebi.ac.uk/interpro/index.html). Comparisons to other fungal genomes can add power to the genomic analysis by providing the evolutionary context of genes. Our goal is to compare the protein coding gene contents of fungal genomes and relate the differences to the physiological differences between species and taxonomic groups. However, no single repository currently provides consistent gene annotations of all sequenced fungi. We have produced consistent Interpro annotations and Tribe-MCL (www.ebi.ac.uk/research/cgg/tribe/) clustering of protein coding sequences of 16 sequenced fungal genomes. Use of clustering allows us to systematically explore undescribed protein families and find members of protein families not yet recognised by Interpro.

Our computational system is based on BioPerl (bio.perl.org) scripts, BioSQL schema for storing the sequences and annotations in a relational database and GBrowse (www.gmod.org) genome browser for visualising genomes, genes and gene clusters. By combining these data sources and technology we can carry out comparative genome analysis and offer a local comparative genome browser for experimental biologists.

We have discovered that the number of genes belonging to protein clusters having members from all fungal species studied is negatively correlated with genome size, e.g. larger fungal genomes are likely to have more specialized functions not present in species with smaller genomes. In contrast based on protein clustering euascomycota and hemiascomycota seem to differ in their level of paralogy. In hemiascomycota average level of paralogy is positively correlated to the size of the genome. In euascomycota, possibly due to Repeat Induced Point mutations, no clear correlation exists. We have also studied in detail differences of protein coding gene content of *Trichoderma reesei* in comparison to other filamentous fungi. *T. reesei* seems to lack many genes related to plant biomass degradation and plant pathogenicity

IIp-9

Transposable element clusters in the genome of *Magnaporthe oryzae* are positively correlated with genetic recombination, loss of synteny, increased gene duplication and increased rate of evolution

Michael R Thon[1*], Thomas K Mitchell[2], Ralph A Dean[2]

[1] Texas A&M University, Department of Plant Pathology & Microbiology, 2132 TAMU, College Station, TX, USA. 77843
[2] North Carolina State University. Center for Integrated Fungal Research, Campus Box 7251 Raleigh, NC, USA. 27695-7251

Transposable elements (TEs) are known to be major contributors to genome evolution. Genome wide studies have shown that in fungi, TEs are usually confined to distinct clusters within the genome. To better understand the role of TEs in mediating genome rearrangement, gene duplication, and gene evolution, we performed an in depth study of chromosome 7 of the rice blast fungus *Magnaporthe oryzae*. Using chromosome 7 as a reference sequence, we identified 21 statistically significant blocks of conserved synteny in *Neurospora crassa*, 17 in *Fusarium graminearum*, and 2 in *Aspergillus nidulans*. In general, the blocks were roughly co-linear and interspersed with intervening, non-syntenic genes. TEs are predominantly restricted to three clusters located in regions that lack conserved synteny. In contradiction to popular evolutionary models as well as observations from other model organism genomes, we found a positive correlation between recombination rate and the distribution of TE clusters on chromosome 7. We grouped chromosome 7 genes into gene families and identified orthologous genes in *N. crassa* and *F. graminearum*. We found that the chromosomal regions defined by the TE clusters have more frequent gene duplications and genes within the clusters are evolving at a faster rate. Together, these data suggest that TEs have a profound impact on the *M. oryzae* genome by creating localized segments with increased rates of chromosomal rearrangements, gene duplications and gene evolution.

The *Leptosphaeria maculans* genome initiative

Marie-Hélène Balesdent[1], Barbara J Howlett[2], Patrick Wincker[3], Thierry Rouxel[1*]

[1] *Phytopathologie et méthodologies de la détection, INRA, F-78026 Versailles, France*
[2] *School of Botany, The University of Melbourne, Parkville, VIC 3010, Australia*
[3] *Genoscope-Centre National de Séquençage, 91057 EVRY Cedex, France*

Genoscope (National Sequencing Centre), the French sequencing agency has been involved in large-scale sequencing of the genome of the Dothideomycete *Leptosphaeria maculans* since 2000. The first two collaborative projects between INRA and Genoscope consisted in precise sequencing, assembly and finishing of a 1.1 Mb genomic region in isolate v23.1.3, along with less precise sequencing of the corresponding region in two other isolates (see L. Gout *et al.* comm.). These data enabled us to characterize the first known retrotransposons and avirulence genes in *L. maculans* (see I. Fudal *et al.*, comm.). They also suggested a particular organisation for the genome of *L. maculans*, encompassing alternating regions of isochores, i.e., of long A+T-rich regions reminiscent of higher Eukaryote heterochromatin, and of G+C-equilibrated, gene-rich regions (see L. Gout *et al.* comm.). With this preliminary snapshot of the genome as a basis, complete shot-gun genome sequencing strategy has been initiated in December 2005. Its goal is to release an annotated assembly with 10-x genome sequence coverage (600 000 reads) for *L. maculans* isolate v23.1.3. Additional sequencing of BAC-ends for physical mapping and ESTs from mycelia grown under three different conditions will also be performed (total of 50 000 reads). The current status of the sequencing project and first data from it will be presented.

IIp-11

Whole genome sequencing of the fungal plant pathogen *Botrytis cinerea*, and preliminary results of the comparison with *Sclerotinia sclerotiorum* genome

Joëlle Amselem[2], François Artiguenave[2], Alain Billault[3], Mathias Choquer[1], Arnaud Couloux[3], Christina Cuomo[4], Martin Dickman[5], Sabine Fillinger[6], Elisabeth Fournier[1], Corinne Giraud[1], Linda Kohn[7], Fabrice Legeai[2], Caroline Levis[1], Cyril Pommier[2], Jean-Marc Pradier[1], Emmanuel Quevillon[2], Béatrice Ségurens[3], Adeline Simon[1], Muriel Viaud[1*], Jean Weissenbach[3], Patrick Wincker[3], Marc-Henri Lebrun[8]

[1] *PMDV, INRA, Rte de St-Cyr, 78026 Versailles, France*
[2] *URGI, INRA, 523 place des Terrasses, 91000 Evry, France*
[3] *Genoscope - CNS, 2 rue Gaston Crémieux, 91000 Evry, France*
[4] *Broad Institute, 320 Charles St., Cambridge MA 02141, USA*
[5] *Inst for Plant Genomics and Biotech, 2123 TAMU, College Station, TX, USA*
[6] *UPMC, INRA, Rte de St-Cyr, 78026 Versailles, France*
[7] *University of Toronto, 3359 Mississauga Rd North, Mississauga ON L5L 1C6, Canada*
[8] *Plant and Fungal Physiology, UMR 2847, CNRS-BayerCropScience, 69009 Lyon, France*

Botrytis cinereais a fungal plant pathogen (leotiomycete) that provokes grey mould on more than 200 dicotyledonous plant species including grapevine. In 2005, the French national sequencing center (Genoscope) has started the genome sequencing of grapevine, *Vitis vinifera*and two of its pathogens: Stolbur phytoplasm and *B. cinerea*. The genome of *B. cinerea*T4 strain (40 Mb) was sequenced with a 10.5 x coverage (about 600 000 reads) and 50.000 ESTs are underway. Genomic sequences (3, 10 and 50 kb inserts) were assembled in 2210 contigs and 107 supercontigs. Structural annotation will be based on automatic gene prediction using *ab initio*(FgenesH) and similarity (genome/cDNA and genome/known proteins comparisons) softwares. Automatic gene prediction will be validated by a manual annotation process involving an international consortium of 20 research groups. In 2005, the Broad Institute released the assembly of a 4-5 x genomic sequence from *B. cinerea*strain B05-10 (TMRI/Syngenta) as well as the 7-8 x genomic sequence of the closely related necrotrophic fungus *Sclerotinia sclerotiorum*(Broad Institute/NFS). Preliminary results on the comparison on their genome organization will be presented.

Comparative analysis of the complete *Pichia stipitis* genome

Andrea Aerts[1*], Asaf Salamov[1], Jane Grimwood[2], Thomas W. Jeffries[3], Igor Grigoriev[1]

[1] DOE Joint Genome Institute, 2800 Mitchell Dr, Walnut Creek, CA 94598, USA.
[2] Stanford Human Genome Center, 975 California Ave, Palo Alto, CA 94304, USA.
[3] USDA Forest Service, Forest Products Laboratory, Madison, WI 53726-2398, USA.

The genome of *Pichia stipitis* is estimated to be approximately 15 million base pairs contained in 8 chromosomes. The genome assembly was annotated using the JGI Annotation Pipeline, which combines various gene prediction, annotation, and analysis tools. Gene models and associated transcripts/proteins are predicted based on cDNA, protein homology and ab initio methods. A gene catalog set is chosen from candidate models based on homology and EST support for intron/exon boundary structure, ORFs and presence of UTR. Finally, each predicted model is analyzed for domain content/structure and functionally annotated. The release v2.0 includes a total of 5841 gene models supported by available EST and cDNA evidence and protein homology. Average gene, transcript and CDS lengths are 1.6kb, 1.5kb and 493 a.a., respectively. Average gene density is 56% with 4204 single exon genes. The genome size, number of genes and CDS lengths are comparable to the numbers found in other sequenced yeast genomes.

A set of 19,635 ESTs was sequenced and clustered at the JGI. 3839 (94%) of the EST clusters mapped to the genome. 2252 (40%) genes are supported by ESTs. An absolute majority of predicted genes are supported by protein homology including 4879 (84%) with strong homology in other fungi (alignment score > 1000). 4083 (70%) of all predicted genes have a predicted protein domain. Manual curation of this genome is on going using the JGI Genome Portal Tools. 2593 genes have been manually curated and there appears to be major differences between *P. stipitis* and other yeasts in oxidative phosphorylation, fatty acid metabolism and fatty acid synthesis.

Pichia stipitis is of fundamental biological interest and important from an applied perspective because it has the highest native capacity for xylose fermentation of any known microbe. We have compared the gene set of *P. stipitis* with the gene sets of five yeasts whose genomes have also been sequenced and assembled. Using comparative methods we have determined a core set of genes common to the six yeasts, as well as the set unique to *P. stipitis*. The results of this comparison attempt to shed light on the metabolic and regulatory networks in native xylose fermenting yeasts in order to make progress in this area.

IIp-13

Genomic comparisons between *Aspergillus flavus* and *A. oryzae* reveal unique genes

Beth L Pritchard[1*], Jiujiang Yu[2], Jennifer Wortman[3], Natalie Fedorova[3], Masayuki Machida[4], Katsuya Gomi[5], Keietsu Abe[5], Ralph A Dean[1], Deepak Bhatnagar[2], Thomas E Cleveland[2], William Nierman[3], Doug E Brown[1], Gary A Payne[1]

[1] North Carolina State University, 840 Main Campus Dr., Raleigh, NC, 27695-7567, USA
[2] USDA/ARS/SRRC, 1100 Robert E. Lee Blvd. New Orleans, LA, 70124, USA
[3] The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD, 20850, USA
[4] Applied Gene Technology Research Group, National Institute of Advanced Industrial Science and Technology (AIST), Central 6, 1-1, Higashi, Tsukuba, Ibaraki, 305-8566, Japan
[5] The, Graduate School of Agricultural Science, Tohoku University, 1-1, Tsutsumidori-Amamiyamachi Aoba-ku, Sendai 981-8555, Japan

Aspergillus flavus and *A. oryzae* are closely related fungi that inhabit very different ecological niches. *A. flavus* is plant and animal pathogen that also produces the toxic and carcinogenic secondary metabolite, aflatoxin. In contrast, *A. oryzae* is the major fungus used in food fermentation and is generally regarded as safe (GRAS). Whole genome sequences are now available for both of these fungi, which allows for a careful comparison of the two species. Overall, these two fungi are very similar in genome size, gene organization and nucleotide identity. The *A. oryzae* genome has been assembled into chromosomes using optical mapping, the high correspondence between the two genomes allowed alignment of the *A. flavus* genomic scaffolds to the *A. oryzae* chromosomes. The 16 largest genomic scaffolds from *A. flavus* essentially correspond to the 16 arms of the 8 predicted chromosomes for *A. oryzae*. However our initial studies show small differences in genome organization due to small-scale insertion-deletion events and transversions, as well as evidence of a translocation event in *A. flavus* between chromosomes II and VI. The translocation break sites and many of the indels are associated with families of uncharacterized repeat elements. Analysis of these repeat elements is ongoing, but amongst those studied to date three putative types of transposable element have been predicted. These transposable elements appear to be present in larger copy numbers in *A. oryzae* than in *A. flavus*. Interestingly, each species has approximately 350 genes unique to that species. Most of the genes are of unknown function, but within this group are genes for secondary metabolism, including polyketide synthases and non-ribosomal peptide synthases. Examination of several of these unique PKS and NRPS genes shows that they are commonly associated with regions of the genome where indels have occurred. This may indicate that there is evolutionary pressure on genes involved in secondary metabolism due to the different ecologies of these fungi. Further analysis of the unique gene sets and comparison of their expression profile will help to identify genes responsible for aflatoxin production and pathogenicity.

Genome sequencing and comparative genomics of fungal pathogens

Christina Cuomo[1], Manfred Grabherr[1], Matthew Rasmussen[2], Chinnappa Kodira[1], Joshua Grochow[2], Evan Mauceli[1], David DeCaprio[1], James Galagan[1], Geraldine Butler[3], Neil Gow[4], Michael Lorenz[5], Sabine Fillinger[6], Marc-Henri Lebrun[7], Jeffery Rollins[8], Linda Kohn[9], Martin Dickman[10], Manolis Kellis[1], Bruce Birren[1]

[1] *Broad Institute of MIT and Harvard, Cambridge, MA, USA*
[2] *Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, USA*
[3] *Conway Institute, University College Dublin, Dublin, IRELAND*
[4] *Institute of Medical Sciences, University of Aberdeen, Aberdeen, UNITED KINGDOM*
[5] *Department of Microbiology, University of Texas Health Science Center, Houston, TX, USA*
[6] *UPMC, INRA, 78026 Versailles, FRANCE*
[7] *Plant and Fungal Physiology, UMR 2847, CNRS-BayerCropScience, 69009 Lyon, FRANCE, Department of Plant Pathology*
[8] *Department of Plant Pathology, University of Florida, Gainesville, FL, USA*
[9] *Department of Botany, University of Toronto, Ontario, CANADA*
[10] *Inst for Plant Genomics and Biotech, TAMU, College Station, TX, USA*

Analysis of an individual fungal genome can be greatly enhanced by comparison to genomes of related species. Comparative genomics highlights sequences conserved between species and can inform such basic questions as how genomes evolve, what genes make a species unique, and what genes and noncoding sequences are shared among species. Two fungal comparative genomics projects under the Broad's Fungal Genome Initiative focus on groups of related fungal pathogens: comparison of the necrotrophic pathogens *Sclerotinia sclerotiorum* and two strains of *Botrytis cinerea*, and the comparison of *Candida* species. *S. sclerotiorum* is an economically devastating fungal pathogen with a broad host range of more than 400 plant species. The 8X assembly, which is 38Mb, was automatically annotated, producing 14,522 predicted genes. We have also generated more than 58,000 EST sequences from three different cDNA libraries representing different developmental stages. A 5X assembly of the Syngenta AG sequence of *B.*cinerea from strain B05.10, which is 39Mb, was subjected to similar automated annotation, producing 16,448 genes. *S. sclerotiorum* and*B.cinerea* share conserved gene order across large regions, and proteins share an average of 79% amino acid identity. A collaborative project is underway to analyze these two genomes and that of the T4 strain of *B. cinerea*. We have also sequenced five related *Candida* species: *C. albicans* (strain WO-1), *C. tropicalis*, *Lodderomyces elongisporus*, *C. guilliermondii*, and *C. lusitaniae*. For the two haploid species, *C. guilliermondii* and *C. lusitaniae*, we have used existing assembly algorithms to generate whole genome draft assemblies. A major challenge in sequencing the other three *Candida* genomes is the difficulty in correctly assembling shotgun sequence from diploid organisms. By modifying our assembly process to represent only one of the two haplotypes in highly polymorphic regions, we released an improved *C. tropicalis* assembly. We have identified preliminary gene sets for each genome assembly, and are refining these by evaluating gene conservation within regions of conserved gene order between species. As part of this process, we are generating comparative genomic resources including whole genome alignments, protein families, species specific genes, and *Candida* specific genes. These resources are being utilized by a community-based comparative analysis project. The growing set of *Candida* genome sequences allows comparisons across a range of evolutionary distances, enabling many different approaches to study the conservation of genes and regulatory elements as well as the evolution of these elements and genomic architecture within *Candida* species.

IIp-15

Genome-wide mutation library in *Aspergillus nidulans*

Jakob Blæsbjerg Nielsen*, Michael Lynge Nielsen, Uffe Hasbro Mortensen

Center for Microbial Biotechnology, BioCentrum-DTU, Technical University of Denmark, Denmark.

For decades the filamentous fungus, *Aspergillus nidulans*, has served as a model organism for numerous industrially and medically significant Aspergilli. One of the main advantages of *A. nidulans* compared to most other Aspergillus species is that it has a sexual cycle, allowing mutant strains to be crossed. The release of the genome sequences for several important Aspergilli has created potential for systematic genome-wide modifications such as gene deletion, promoter replacements, fluorescent protein tags and allele replacements. We have recently developed an efficient PCR based genome manipulation system for *A. nidulans*, which improves targeted integration of exogenous DNA while reducing ectopic integration. Inspired by the success of the gene deletion library from *Saccharomyces cerevisiae*, we intend to use this technology to initiate a genome-wide mutation library in *A. nidulans*. We here present our progress in the improvement and scale up of our current gene targeting capabilities, with the aim of providing a high-throughput genome-wide manipulation system that will be transferable to other fungal species.

Genetic linkage map and expression analysis of genes expressed in the lamellae of the edible basidiomycete *Pleurotus ostreatus*

Antonio G. Pisabarro, Sang-Kyu Park, María M. Peñas, Lucía Ramírez

Department of Agrarian Production, Public University of Navarre, 31006 Pamplona, Spain

Pleurotus ostreatus is an industrially cultivated basidiomycete with nutritional and environmental applications. Its genome contains 35 Mbp organized in 11 chromosomes. There is currently available a genetic linkage map based predominantly on anonymous molecular markers complemented with the mapping of QTLs controlling growth rate and industrial productivity. In order to increase the saturation of the existing linkage maps, we have identified and mapped 82 genes expressed in the lamellae. Their manual annotation revealed that 34.1 % of the lamellae-expressed and 71.5 % of the lamellae-specific genes correspond to previously unknown sequences or to hypothetical proteins without a clearly established function. Furthermore, the expression pattern of some genes provides an experimental basis for studying gene regulation during the change from vegetative to reproductive growth. Finally, the identification of various differentially regulated genes involved in protein metabolism suggests the relevance of these processes in fruit body formation and maturation.