



# **II Comparative Fungal Genomics**

**Chair:**

**Merja Penttilä & Gerhard Adam**

## Ilo-1

### The *Podospora* genome project

Genoscope<sup>2</sup>, Consortium Podospora<sup>3</sup>, Fabienne Malagnac<sup>1</sup>

<sup>1</sup> Institut de Genetique et Microbiologie, UMR8621, Orsay, France

<sup>2</sup> Patrick Wincker, Arnaud Couloux, Corinne Dasilva, Jean Weissenbach. Genoscope, Evry, France

<sup>3</sup> Philippe Silar, Olivier Lespinet, Eric Espagne, Hamid Khalili, Robert Debuchy, Sylvie Arnaise, Veronique Berteaux-Lecellier, Veronique Contamine, Evelyne Coppin, Michelle Dequard-Chablat, Marguerite Picard: Institut de Genetique et Microbiologie, UMR8621, Orsay, France. Antoine Boivin, Carole H. Sellem, Annie Sainsard-Chanet: Centre de Genetique Moleculaire, UPR2167, Gif sur Yvette, France. Corinne Clave, Bérengere Pinan-Lucarre, Sven Saupe: Institut de Biochimie et Genetique Cellulaires, UMR5095, Bordeaux, France. Fons Debets, Rolf Hoekstra: Laboratory of Genetics, Wageningen University, Wageningen, Netherland

Using a whole genome shotgun approach, we sequenced the *Podospora anserina* genome to a 10X coverage. Arachne assembly led to 1891 contigs, grouped in 192 supercontigs totaling 35 Mbp, close to pulsed-field gel estimate of genome size. 36 supercontigs account for 98% of the physical map. Previously identified genetic markers and 150 newly discovered microsatellites allowed supercontigs to be anchored to the genetic map. 6 000 ESTs, representing more than 50% of the total of 11 000 predicted genes, obtained from various stages of *Podospora* life cycle had been sequenced in order to be used during the currently on-going annotation process. Data are posted at <http://podospora.igmors.u-psud.fr>.

## Community Involved Fungal Genome Annotation at the JGI-LANL

Gary Xie<sup>1\*</sup>, Randy M Berka<sup>2</sup>, Bernard Henrissat<sup>3</sup>, Markku Saloheimo<sup>4</sup>, Jon Magnuson<sup>5</sup>, Alfredo Lopez De Leon<sup>2</sup>, Daniel Cullen<sup>6</sup>, Beth Nelson<sup>2</sup>, Mikko Arvas<sup>4</sup>, Ann Westerholm-Parvinen<sup>4</sup>, Sandy Merino<sup>2</sup>, Scott E Baker<sup>5</sup>, Monika Schmoll<sup>7</sup>, Paul Harris<sup>2</sup>, Mick Ward<sup>8</sup>, Luis F Larrondo<sup>9</sup>, Jian Yao<sup>8</sup>, Merja Penttilä<sup>4</sup>, Melissa Jackson<sup>1</sup>, Monica Misra<sup>1</sup>, David Goodstein<sup>11</sup>, Nina Thayer<sup>1</sup>, Astrid Terry<sup>11</sup>, Wayne Huang<sup>11</sup>, Sam Rash<sup>11</sup>, Paramvir Dehal<sup>11</sup>, Nick Putnam<sup>11</sup>, Jarrod Chapman<sup>11</sup>, Barbara Robbertse<sup>10</sup>, Christian Kubicek<sup>7</sup>, Dan Rokhsar<sup>11</sup>, Paul Gilna<sup>1</sup>, Tom Brettin<sup>1</sup>, Susan Lucas<sup>11</sup>, Diego A Martinez<sup>1</sup>

<sup>1</sup> DOE Joint Genome Institute (JGI) at Los Alamos National Laboratory

<sup>2</sup> Novozymes, Inc.

<sup>3</sup> CNRS and Universités Aix-Marseille I & II

<sup>4</sup> VTT Biotechnology

<sup>5</sup> Pacific Northwest National Laboratory

<sup>6</sup> USDA Forest Products Lab at University of Wisconsin-Madison

<sup>7</sup> TU Vienna Institute of chemical engineering

<sup>8</sup> Genencor International

<sup>9</sup> Dartmouth Medical School

<sup>10</sup> Cornell University

<sup>11</sup> DOE Joint Genome Institute (JGI) at Walnut Creek

The role of the fungal genome annotation team at LANL is to provide high quality manually curated annotation of fungal genomic sequence. The JGI-LANL and fungal research community are teaming up to publish a full analysis and annotation of *Trichoderma reesei* and *Aspergillus niger* chromosomes. Work on the *T. reesei* genome was completed earlier this year and will appear in publication in the coming months. We begin our analysis by running each genome through an automated pipeline that combines evidence-based and ab-initio methods. All information is displayed before manual review in the JGI's Genome Portal, available at <http://genome.jgi.doe.gov/>. Utilizing a combination of the Genome Portal, local tools and the Apollo editing system each predicted gene is then manually reviewed by comparing the predicted gene to EST and mRNA sequences from the *T. reesei* and *A. niger* genomes as well as closely related species, such as *Neurospora crassa*, *Fusarium graminearum*, *Magnaporthe grisea* and other *Apergillus* species. This suite of tools allows a distributed group of annotators from the fungal research community to correct errors as necessary and enrich the automated information with comments and human annotations. All manual annotations are immediately viewable and searchable in the Genome Portal.

## 110-3

### ***Aspergillus flavus* genome sequence and comparative analysis with *Aspergillus oryzae***

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

<sup>1</sup> North Carolina State University, 840 Main Campus Dr., Raleigh, NC, 27695-7567, USA

<sup>2</sup> USDA/ARS/SRRC, 1100 Robert E. Lee Blvd. New Orleans, LA, 70124, USA

<sup>3</sup> The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD, 20850, USA

<sup>4</sup> Applied Gene Technology Research Group, National Institute of Advanced Industrial Science and Technology (AIST), Central 6, 1-1, Higashi, Tsukuba, Ibaraki, 305-8566, Japan

<sup>5</sup> The Graduate School of Agricultural Science, Tohoku University, 1-1, Tsutsumidori-Amamiyamachi Aoba-ku, Sendai 981-8555, Japan

*Aspergillus flavus* is a filamentous fungus pathogenic to plants and animals. It attacks several crop species and produces the potent carcinogen, aflatoxin, which contaminates agricultural grains. *Aspergillus flavus* is also the second leading cause of aspergillosis in humans. A whole genome sequencing project funded by the USDA/NRI and USDA/ARS has been completed at The Institute for Genomic Research (TIGR). The 5X sequence has been assembled from 2080 contigs into 79 scaffolds. Scaffold sizes range from 4.5 Mb to 1.0 kb, and 75% of the genome is represented in the 10 largest scaffolds. Automated gene finding done at TIGR predicts that the 36.3 Mb genome contains 13,071 genes. Our analyses show that 99.6% of these predicted genes reside in the 16 largest scaffolds. The sequence data has been deposited at the NCBI GenBank Database (<http://www.ncbi.nlm.nih.gov>) and is also available at the website (<http://www.aspergillusflavus.org>) through a web browser that allows visualization of Blast matches to genes, proteins and genomic sequence of other *Aspergillus* species, alignments of *A. flavus* ESTs, and GO annotations. Fortunately, a whole genome sequence is also available for *A. oryzae*, a closely related fungus that does not produce aflatoxins and has lost the ability to infect plants, animals and humans. Comparative genomic analysis between the two species indicates that the genome of *A. flavus* is very similar to that of *A. oryzae*, sharing greater than 96% sequence identity. Further, these two fungi have a similar number of genes necessary for secondary metabolism. The predicted number of genes for polyketide synthases (34), non-ribosomal peptide synthases (22), and cytochrome P450s (122) is similar to that predicted for *A. oryzae*, 31, 24, and 151, respectively. In contrast, *A. nidulans* is predicted to have 14 polyketide synthases, 13 non-ribosomal peptide synthases and 65 cytochrome P450s. An optical map of the *A. oryzae* genome has allowed direct comparison of the physical structure of these two organisms. The high degree of DNA similarity between these two fungi allowed alignment of the *A. flavus* genomic scaffolds to the *A. oryzae* chromosomes predicted by the optical map. The 16 largest genomic scaffolds from *A. flavus* essentially correspond to the 16 arms of the 8 predicted chromosomes for *A. oryzae*. Whole genome Affymetrix arrays will be available for *A. flavus* soon, allowing an analysis of differences in gene transcription between these two fungi.

## Sequencing *Mycosphaerella* and *Cercospora* species will revolutionize the control of major global threats on wheat, banana and maize

Stephen B Goodwin<sup>1\*</sup>, Larry D Dunkle<sup>1</sup>, Alice C. L. Churchill<sup>2</sup>, Jean Carlier<sup>3</sup>, Andy James<sup>4</sup>, Manoel Souza<sup>5</sup>, Pedro Crous<sup>6</sup>, Nicolas Roux<sup>7</sup>, Theo A. J. van der Lee<sup>8</sup>, Cees Waalwijk<sup>8</sup>, Erika Lindquist<sup>9</sup>, Jim Bristow<sup>9</sup>, Gert H. J. Kema<sup>8</sup>

<sup>1</sup> USDA-ARS/Purdue University, 915 West State Street, West Lafayette, IN 47907-2054, USA

<sup>2</sup> Department of Plant Pathology, Cornell University, Ithaca, NY 14853, USA

<sup>3</sup> UMR BGPI, CIRAD, Montpellier, France

<sup>4</sup> CICY, Merida, Mexico

<sup>5</sup> EMBRAPA Genetic Resources & Biotechnology, Brasilia, Brazil

<sup>6</sup> Fungal Biodiversity Centre, Utrecht, The Netherlands

<sup>7</sup> INIBAP, Montpellier, France

<sup>8</sup> Plant Research International B.V., Wageningen, The Netherlands

<sup>9</sup> DOE-Joint Genome Institute, Walnut Creek, CA, USA

*Mycosphaerella* is one of the largest genera of plant pathogenic fungi with more than 1,000 named species, many of which are important pathogens causing leaf spotting diseases in cereals, citrus, banana, and eucalypts, soft fruits such as strawberry, and horticultural crops including many *Brassica* species. A few species of *Mycosphaerella* cause disease in humans and other vertebrates. The major *Mycosphaerella* plant pathogens include *M. graminicola* of wheat and *M. fijiensis* of banana, which require global annual fungicide inputs of \$400 million and \$2.5 billion, respectively. Both fungi are very important to the world economy. For example, *M. graminicola* causes more than \$275 million in losses annually to U.S. wheat growers, and more than 70% of the fungicides sprayed on wheat in Europe are to combat this pathogen. The International *Mycosphaerella* Genomics Consortium was started with the goal to sequence the genomes of *M. graminicola* and *M. fijiensis*. The former was chosen due to its genetic tractability, availability of extensive genomic resources, large worldwide research community and phylogenetic distinctiveness from other fungi being sequenced. The latter was chosen because of its worldwide impact on banana production. A joint project between the USDA-ARS/Purdue University and Plant Research International B.V. was initiated to sequence both genomes, along with 40,000 ESTs from both *M. fijiensis* and the related maize pathogen *Cercospora zea-maydis*. The work was conducted through the Community Sequencing Program sponsored by the U.S. DOE-Joint Genome Institute. The initial goals of the projects are to: assemble 8x genomic shotgun sequences of *M. graminicola* strain IPO323 and *M. fijiensis* strain CIRAD 86; perform automated annotations of these genomic sequences and directed annotations using the ~80,000 ESTs from both *Mycosphaerella* species (the *M. graminicola* set will be made available by Syngenta); and make these sequences available publicly through a series of linked web sites for comparative analyses. Currently, the 8x *M. graminicola* sequence (518,271 traces) is available at NCBI. An initial assembly was made after the 4x sequencing was complete. The 289,742 reads were organized into 2962 contigs spanning 37.65 Mb. These contigs were assembled into 187 scaffolds covering 39.05 Mb, giving a revised genome size of 41.8 Mb, slightly larger than estimated previously. In addition, a draft mitochondrial assembly yielded a 43,962-base scaffold that appears to cover the complete genome. A community-wide annotation effort culminating in an annotation jamboree is anticipated for later during 2006 and will be open to all interested participants.

## IIo-5

### **Comparative organisation of the *AvrLm1–AvrLm6* genomic environment within the *Leptosphaeria maculans* – *L. biglobosa* species complex and in *Stagonospora nodorum***

**Maria R. Eckert<sup>1</sup>, Lilian Gout<sup>1</sup>, Simon Ross<sup>1</sup>, Isabelle Fudal<sup>1</sup>, Bronislava Profotova<sup>1</sup>, Laurence Cattolico<sup>2</sup>, Sylvie Bernard-Samain<sup>2</sup>, Thierry Rouxel<sup>1\*</sup>, Marie-Hélène Balesdent<sup>1</sup>**

<sup>1</sup> INRA-PMDV, Route de Saint Cyr, 78026 Versailles Cedex, France

<sup>2</sup> Genoscope, Centre National de Séquençage, 91057 Evry Cedex, France

Stem canker of oilseed rape is caused by two closely associated *Leptosphaeria* species, *L. maculans* and *L. biglobosa*, which can be further divided into seven sub-species, depending on their host plant specificity and geographic distribution. These phytopathogenic Ascomycetes belong to the Dothideomycetes class, for which only one representative, *Stagonospora nodorum* (formerly known as *Leptosphaeria nodorum*) has been fully sequenced to date ( [www.broad.mit.edu/annotation/fungi/stagonospora\\_nodorum](http://www.broad.mit.edu/annotation/fungi/stagonospora_nodorum) ).

In a collaborative project between GENOSCOPE and INRA we recently sequenced 1.1 Mb of the *AvrLm1 – AvrLm6* avirulence gene cluster genomic environment of a *L. maculans* 'brassicae' isolate. The region is characterised by 4 small (20 – 70kb) GC-rich, gene-rich regions interspersed by large AT-rich, non-coding regions.

Occurrence of ortholog genes and conservation of their organisation was assessed within the *L. maculans*-*L. biglobosa* species complex and within the closely related *S. nodorum* using (i) PCR amplification of three genes from each of the 4 GC-rich regions, (ii) hybridisation of the PCR product on digested genomic DNA and electrokaryotypes and (iii) comparison of the 1.1 Mb *L. maculans* sequence with the whole *S. nodorum* genome sequence.

Using these approaches we found that most of the genes in the 1.1 Mb *L. maculans* region were strongly conserved in *S. nodorum* and located to the same genomic region, scaffold 1.7 (40 genes, 71.4%), with numerous small-scale breaks of synteny. Five (41.7%) of the genes investigated by PCR/hybridisation were conserved within the species complex and organized similarly in all 7 sub-species, whereas 5 (41.7%) of the genes were conserved within *L. maculans* subspecies only, but not *L. biglobosa*. Two genes (16.7%) were specific to the *L. maculans* 'brassicae' and did not amplify DNA from the closely related 'lepidii' sub-species. The conserved genes are involved in primary metabolism whereas the genes specific to *L. maculans* are secondary metabolism genes and/or genes putatively involved in fungal pathogenicity.

## **Omics technologies, an invitation to use these novel tools for improving the *A. niger* cell factory.**

**Cees Sagt<sup>\*</sup>, Noël van Peij, Maurien Olsthoorn, Rogier Meulenberg, Hein Stam, Han de Winde, Herman Pel**

*DSM Food Specialties R&D/Genetics P.O. Box 1, 2600 MA Delft The Netherlands*

The genome of the industrially important filamentous fungus *Aspergillus niger* was sequenced and annotated. Moreover Affymetrix GeneChip<sup>®</sup> arrays were designed and used for full transcriptomics studies in *A. niger*. In addition we have developed a proteomics approach for *A. niger* which is aimed at studying genome wide protein expression, both extracellular and intracellular. This integrated “omics” approach allowed us to gain insight in physiological processes related to protein production by the cell factory *A. niger*.

Cells were grown under controlled conditions in fed-batch fermentations. Samples were studied on transcript level as well as protein level. These omics results were integrated and interpreted in order to gain insight in cellular processes of *A. niger* in down-scaled industrial fermentations. Data analysis was primarily done by a data driven approach, followed by a knowledge based selection and interpretation. We will present our integrated way of working based on the use of omics technologies.

The implementation of omics technologies in industrial R&D has revolutionized the process of strain and process improvement. The impact of this revolution will be discussed as well as potential pitfalls. In addition we will present a model containing the most important aspects of the *A. niger* cell factory identified by genome annotation, transcriptomics and proteomics. The academic community is now able to benefit from available omics tools which will accelerate further fungal research. This synergy between industrial R&D and academic research should be a key issue for the future.