

Cambridge University Press

978-0-521-85935-6 - Exploitation of Fungi

Edited by G. D. Robson, P. van West and G. M. Gadd

Excerpt

[More information](#)

---

# I

## **Comparative and functional fungal genomics**

Cambridge University Press

978-0-521-85935-6 - Exploitation of Fungi

Edited by G. D. Robson, P. van West and G. M. Gadd

Excerpt

[More information](#)

## 1

## Genome and proteome analysis of industrial fungi

S. E. BAKER, C. F. WEND, D. MARTINEZ,  
J. K. MAGNUSON, E. A. PANISKO, Z. DAI,  
K. S. BRUNO, K. K. ANDERSON, M. E. MONROE,  
D. S. DALY AND L. L. LASURE

### Background

In order to decrease dependence on petroleum, the United States Department of Energy (USDOE) Office of the Biomass Program (OBP) is investing in research and development to enable its vision of the biorefinery. The biorefinery will decrease the use of petroleum through conversion of biomass such as crops or agricultural waste into fuels and products.

In 2004, the USDOE OBP asked researchers at the Pacific Northwest National Laboratory (PNNL) and the National Renewable Energy Laboratory (NREL) to prepare a list of the top ten building-block chemicals that can be derived from simple sugars by biological and/or chemical means. The resulting list of twelve building-block chemicals and the accompanying report ([www.eere.energy.gov/biomass/pdfs/35523.pdf](http://www.eere.energy.gov/biomass/pdfs/35523.pdf)) form an informational foundation on which future DOE and industry bioproducts research will be built (Table 1.1).

How do fungi fit into the biorefinery? Analysis of the ‘top ten’ study indicates that nine of the top twelve chemical building blocks are currently produced, or may potentially be produced, by fungal fermentation processes. However, a significant barrier to the use of bio-based products is the economic feasibility – fuels and products must be price-competitive with those derived from petroleum. An obvious way to decrease the costs of biobased products from fungi is to make fermentation strains more productive and processes more efficient. Traditional strain improvement programmes typically span a timescale measured in decades and process development done through the use of batch cultures is extremely labour intensive.

How can current and future bioproducts mycology research in this area take advantage of the growing list of fungi whose genomes have been

*Exploitation of Fungi*, eds. G. D. Robson, P. van West and G. M. Gadd. Published by Cambridge University Press. © Cambridge University Press 2007.

Table 1.1 *Top 12 sugar-derived chemical building blocks (potentially derived from fungi in bold).*

<b>1, 4 succinic, fumaric and malic acids</b>
2, 5 furan dicarboxylic acid
<b>3 hydroxy propionic acid</b>
<b>aspartic acid</b>
<b>glucaric acid</b>
<b>glutamic acid</b>
<b>itaconic acid</b>
levulinic acid
3-hydroxybutyrolactone
<b>glycerol</b>
<b>sorbitol</b>
<b>xylitol/arabinitol</b>

sequenced, as well as the cutting edge systems biology methods in order to bring the vision of the biorefinery to reality? With genome sequences available for several fungi of industrial importance and the ability to grow fungi at a steady state, the prospects for decreased time and cost investments for economical fungal bioprocesses that produce fuels and products have become considerably brighter.

**What can fungal genomics offer industrial mycology?**

At its most basic level, the genome sequence of an organism provides a high-level view of the biology of an organism. The presence or absence of genes or gene families can give glimpses into the lifestyle and ecological niches of a given organism. For example, the presence of genes that are predicted to be involved in sexual development in the ‘asexual’ opportunistic pathogen, *Aspergillus fumigatus*, has led researchers to hypothesize that this organism does have sex (Paoletti *et al.*, 2005). Since current or potential commodity chemicals found in the ‘top ten’ list are central metabolites, it is valuable to understand the central metabolic pathways present or deficient in a particular fungus of interest. Also, since the goal of the biorefinery is to use complex biomass sources of carbon (polysaccharides) it is important to understand the suite of plant polymer degrading enzymes present in the genome of an industrially relevant fungus in order to ascertain the substrate range of the organism. The genomic sequence from an organism serves as the foundation for further study of the organism at the global scale. In particular,

Cambridge University Press

978-0-521-85935-6 - Exploitation of Fungi

Edited by G. D. Robson, P. van West and G. M. Gadd

Excerpt

[More information](#)

transcriptomics and proteomics methodologies are greatly enhanced by the presence of genomic sequence and predicted protein databases. Furthermore, encoded in the genome are all of the metabolic pathways that encode the ability of fungi to make an amazingly diverse array of compounds. The challenge that confronts the fungal research community at present is how to make use of all of the genomic information to further as many avenues of mycological research as possible.

#### The *Aspergillus niger* genome project

By the end of 2005, three different strains of *Aspergillus niger* will have been sequenced by three different organizations. Two strains have been sequenced by industry. A US company, Integrated Genomics, Inc., sequenced ATCC strain 9029, a wild-type strain that can produce citric acid. The sequence coverage for this strain is low (4–6X) and distributed across over 9000 contigs. The sequence was purchased by Pacific Northwest National Laboratory in 2004 and is currently available upon request. A derivative of the *A. niger* strain NRRL 3122 was sequenced by a Dutch company, DSM.

Most recently, the USDOE Joint Genome Institute (JGI) in collaboration with the PNNL Fungal Biotechnology Team began the first public *A. niger* sequencing project. ATCC strain 1015 was chosen as the target strain for sequencing. This choice was attractive for sequencing for a number of reasons. From a historical point of view, this strain was the first used for a patented citric acid process in 1917 (Currie, 1917). Furthermore, a derivative of the strain exists (ATCC strain 11414) that has been used for studies of citric acid production and fermentation morphology control (Perlman, Dorrell & Johnson, 1946; Shu & Johnson, 1947; Dai *et al.*, 2004). A genomic sequence and automated gene models are expected to be available through the JGI Genome Portal by early 2006.

#### The *Trichoderma reesei* genome project

*Trichoderma reesei*, the asexual anamorph of *Hypocrea jecorina* (Kuhls *et al.*, 1996), is one of the most prodigious generators of protein known. *T. reesei* was discovered in the south Pacific island of Samoa, degrading the canvas tents used by the US forces during World War II. The particular family of proteins for which *T. reesei* is most studied, cellulases (enzymes that break down cellulose), were the cause of the undesirable tent degrading capability of this ascomycete fungus. *T. reesei* has since become an important organism in biotechnology for the production of the now desirable cellulases and other valuable proteins.

Cambridge University Press

978-0-521-85935-6 - Exploitation of Fungi

Edited by G. D. Robson, P. van West and G. M. Gadd

Excerpt

[More information](#)6 *S. E. Baker et al.*

The USDOE began working with two biotechnology companies, Genencor and Novozymes, to bring down the cost of cellulases so that they could be economically applied in the conversion of lignocellulosic material to simple sugars and then to ethanol or chemical products as alternatives to petroleum based fuels and products. To assist this effort, the JGI began sequencing the genome in 2002. The genome of *T. reesei* and the planned sequencing of two close relatives of *T. reesei*, *Trichoderma virens* and *Trichoderma atroviride*, both by the JGI, will allow researchers to use powerful comparative genomics techniques to help understand the complicated interplay of genes and regulatory elements involved in the production of important proteins.

As in all whole genome level projects, understanding the mass of data generated requires bioinformatics tools. To meet this need, the JGI has released version 1 of the genome to the public in an interactive portal available on the internet (<http://genome.jgi-psf.org>). Currently, the JGI is continuing to finish the genome at its Los Alamos National Laboratory site and will release version 2 to the public in early 2006. Version 2, containing updated annotations, will consist of fewer than 100 scaffolds, a ten-fold reduction over version 1.

### **Fungal proteomics**

How can genome information be leveraged to gain high throughput global functional data? The rapidly developing field of proteomics – the study of the protein complement of an organism – has the potential to revolutionize how we study biological systems. Proteomic methodologies are being used in the study of the environment, as medical diagnostics, and in basic and industrial research settings (for example, Sadygov, Cociorva & Yates, 2004; Jacobs *et al.*, 2005; Ram *et al.*, 2005). One important issue for proteomic methodologies is the quality of the data, i.e. are we really seeing what we think we are seeing?

Current proteomics methodologies using both two-dimensional gel electrophoresis and liquid chromatography/mass spectrometry allow a snapshot of the proteins expressed by an organism at the particular point in time when the organism was processed for protein purification and analysis. The large amounts of data generated by ‘shotgun’ proteomic analysis often serve as the foundation for more detailed, hypothesis-driven research on individual proteins or groups of proteins. Increased quality of proteomics data will have dramatically positive effects on subsequent studies that use the data. Through the use of various statistical analysis techniques and internal protein/peptide standards we have

Cambridge University Press

978-0-521-85935-6 - Exploitation of Fungi

Edited by G. D. Robson, P. van West and G. M. Gadd

Excerpt

[More information](#)

been able to improve fundamentally the way that proteomic data is analyzed.

Current work at PNNL is focused on developing statistical analysis tools that allow for relative quantification of proteins between different samples. The statistical methodology casts a new physical description of measured peptide abundances as a mixed effects linear statistical model, and uses restricted maximum likelihood estimation (REML) to estimate model parameters and associated errors.

### **The case for filamentous fungal chemostat research in process development**

In bioprocess research laboratories, development of a new process has traditionally employed iterative experiments consisting of culturing fungi in shake flasks, small fermentors and process prototype fermentors. One of the significant bottlenecks in this standard approach is the requirement for large numbers of repetitive experiments using batch fermentors. This is an inherently limiting step due to the amount of labour involved, equipment availability and that typically only a single variable can be examined per experiment. Large numbers of shake-flask experiments are often performed previous to or in parallel with the batch fermentor experiments to examine more nutritional and physical process variables. However, shake flasks are not amenable to process control and monitoring for dissolved oxygen, pH and other variables. An alternative to this standard approach is to use continuous culture technology, often referred to as chemostat or auxostat technology.

To date, relatively little chemostat research has been performed with filamentous fungi. Examples include fermentations to obtain various types of information, from examination of catabolic carbon flow, to stability determination of gene copies (Ng, Smith & McIntosh, 1973; Kristiansen & Sinclair, 1979; Wiebe & Trinci, 1991; Gallmetzer & Burgstaller, 2002). Some investigations have been performed in auxostats to evaluate maximum growth or product rates based upon a single limiting parameter (Simpson *et al.*, 1995; Swift *et al.*, 1998). Chemostat systems have been examined where clumps or flocs of filamentous morphology occur under certain agitation or flow rates (Amanullah, Tuttiett & Nienow, 1998).

For accelerating the rate of progress of investigation and optimization of novel bioprocesses continuous culture technology will be crucial. For example, growth rate and product formation rate can be measured in response to several carbon sources or different pH levels in a single continuous culture experiment. The medium composition can be optimized in

an experiment by examining major (carbon source, nitrogen source) and minor (salts, micronutrients) medium components. The responses of the growth and product formation rate (efficiency) to critical parameters can be plotted and analyzed by statistical methods to arrive at the optimum. Translating this steady-state data back into the batch system where growth and production rate vary is a challenge to be addressed. This will be important for proving the utility of the method to industrial processes where the final process is nearly always operated in a batch mode on large scales.

An additional advantage of chemostat technology is the improvement of uniformity and reproducibility of biological samples. Proteomics research will benefit greatly from improvements in biological sample quality since this directly impacts the quality of the resulting proteomics data. A fungus grown in a chemostat can be examined under conditions of steady-state growth. Once steady-state growth is established, critical factors that may affect an organism's physiological state and therefore its proteome can be assessed by sampling, changing a parameter, re-establishing steady state and sampling again. Finally, chemostat technology will be useful in applying the principles of quantitative genetics to the improvement of desired characteristics in a target fungus through the use of selective pressure.

### Conclusions

A systems biology approach is needed to utilize fully the tools available for the study of all fungi with potential roles to play in the biorefinery. While one challenge for industrial mycology lies in the development of the strains to supply the biorefinery with enzymes and products, another will be the integration of huge fungal organism datasets generated by genomics, proteomics and bioprocessing technology. Indeed, connecting metabolite and growth kinetic data from exploration of bioprocess parameter space with proteome and genome data should be a priority for this field.

A recent study funded by the USDOE OBP indicates that a number of building-block compounds derived from sugars that can feed into industrial chemical processes have the potential to be made through fungal fermentation processes. While central metabolism for the generic organism has been described, detailed descriptive and functional information for specific organisms is not well known. A combination of bioinformatics, bioprocess and proteomic experimentation and analysis for each fungus for which there is genomic information will be a powerful tool as we model

Cambridge University Press

978-0-521-85935-6 - Exploitation of Fungi

Edited by G. D. Robson, P. van West and G. M. Gadd

Excerpt

[More information](#)

how genetic alteration and growth parameter manipulation for improved fungal strains affects enzyme and organic acid production.

## References

- Amanullah, A., Tuttiett, B. & Nienow, A. W. (1998). Agitator speed and dissolved oxygen effects in xanthan fermentations. *Biotechnology and Bioengineering*, **57**, 198–210.
- Currie, J. N. (1917). The citric acid fermentation of *Aspergillus niger*. *Journal of Biological Chemistry*, **31**, 15–37.
- Dai, Z., Mao, X., Magnuson, J. K. & Lasure, L. L. (2004). Identification of genes associated with morphology in *Aspergillus niger* by using suppression subtractive hybridization. *Applied Environmental Microbiology*, **70**, 2474–85.
- Gallmetzer, M. & Burgstaller, W. (2002). Efflux of organic acids in *Penicillium simplissimum* is an energy-spilling process, adjusting the catabolic carbon flow to the nutrient supply and the activity of catabolic pathways. *Microbiology*, **148**, 1143–9.
- Jacobs, J. M., Adkins, J. N., Qian, W. J., Liu, T., Shen, Y., Camp, D. G. 2nd & Smith, R. D. (2005). Utilizing human blood plasma for proteomic biomarker discovery. *Journal of Proteome Research*, **4**, 1073–85.
- Kristiansen, B. & Sinclair, C. G. (1979). Production of citric acid in continuous culture. *Biotechnology and Bioengineering*, **21**, 297–315.
- Kuhls, K., Lieckfeldt, E., Samuels, G. J., Kovacs, W., Meyer, W., Petrini, O., Gams, W., Borner, T. & Kubicek, C. P. (1996). Molecular evidence that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete *Hypocrea jecorina*. *Proceedings of the National Academy of Sciences USA*, **93**, 7755–60.
- Ng, A. M. L., Smith, J. E. & McIntosh, A. F. (1973). Conidiation of *Aspergillus niger* in continuous culture. *Archives of Microbiology*, **88**, 119–26.
- Paoletti, M., Rydholm, C., Schwier, E. U., Anderson, M. J., Szakacs, G., Lutzoni, F., Debeaupuis, J. P., Latge, J. P., Denning, D. W. & Dyer, P. S. (2005). Evidence for sexuality in the opportunistic fungal pathogen *Aspergillus fumigatus*. *Current Biology*, **15**, 1242–8.
- Perlman, D., Dorrell, W. W. & Johnson, M. J. (1946). Effect of metallic ions on the production of citric acid by *Aspergillus niger*. *Archives of Biochemistry*, **11**, 131–43.
- Ram, R. J., Verberkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., Blake, R. C. 2nd, Shah, M., Hettich, R. L. & Banfield, J. F. (2005). Community proteomics of a natural microbial biofilm. *Science*, **308**, 1915–20.
- Sadygov, R. G., Cociorva, D. & Yates, J. R. 3rd. (2004). Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. *Nature Methods*, **1**, 195–202.
- Shu, P. & Johnson, M. J. (1947). Effect of the composition of the sporulation medium on citric acid production by *Aspergillus niger* in submerged culture. *Journal of Bacteriology*, **54**, 161–7.
- Simpson, D. R., Wiebe M. G., Robson G. D. & Trinci A. P. J. (1995). Use of pH auxostats to grow filamentous fungi in continuous-flow culture at maximum specific growth-rate. *FEMS Microbiology Letters*, **126**, 151–7.
- Swift, R. J., Wiebe, M. G., Robson, G. D. & Trinci A. P. J. (1998). Recombinant glucoamylase production by *Aspergillus niger* B1 in chemostat and pH auxostat cultures. *Fungal Genetics and Biology*, **25**, 100–9.
- Wiebe, M. G. & Trinci A. P. J. (1991). Dilution rate as a determinant of mycelial morphology in continuous culture. *Biotechnology and Bioengineering*, **38**, 75–81.



Cambridge University Press

978-0-521-85935-6 - Exploitation of Fungi

Edited by G. D. Robson, P. van West and G. M. Gadd

Excerpt

[More information](#)

## 2

## The rice blast story: from genome sequence to function

R. A. DEAN, T. MITCHELL, R. KULKARNI,  
N. DONOFRIO, A. POWELL, Y. Y. OH, S. DIENER,  
H. PAN, D. BROWN, J. DENG, I. CARBONE,  
D. J. EBBOLE, M. THON, M. L. FARMAN,  
M. J. ORBACH, C. SODERLUND, J.-R. XU,  
Y.-H. LEE, N. J. TALBOT, S. COUGHLAN,  
J. E. GALAGAN AND B. W. BIRREN

### Introduction

Rice blast disease, caused by the filamentous fungus *Magnaporthe grisea*, is a serious and recurrent problem in all rice-growing regions of the world (Talbot, 2003; Valent & Chumley, 1991). It is estimated that each year enough rice is destroyed by rice blast disease to feed 60 million people. Control of this disease is difficult; new host-specific forms develop quickly to overcome host resistance and chemical control is typically not cost effective (Ou, 1987). Infections occur when fungal spores land and attach themselves to leaves using a special adhesive released from the tip of each spore (Hamer *et al.*, 1988). The germinating spore develops an appressorium, a specialized infection cell, which generates enormous turgor pressure – up to 8 MPa – that ruptures the leaf cuticle allowing invasion of the underlying leaf tissue (de Jong *et al.*, 1997; Dean, 1997). Subsequent colonization of the leaf produces disease lesions from which the fungus sporulates and spreads to new plants. When rice blast infects young rice seedlings, whole plants often die, while spread of the disease to the stems, nodes or panicle of older plants results in nearly total loss of the rice grain. Recent reports have further shown that the fungus has the capacity to infect plant roots (Sesma & Osbourn, 2004). Different host-limited forms of *Magnaporthe* also infect a broad range of grass species including wheat, barley and millet.

The rice blast fungus has emerged as a seminal model to elucidate the molecular basis of infection-related development and host specificity (Valent & Chumley, 1991). The fungus is haploid, has a relatively small genome contained in seven chromosomes, can be induced to form

*Exploitation of Fungi*, eds. G. D. Robson, P. van West and G. M. Gadd. Published by Cambridge University Press. © Cambridge University Press 2007.

Cambridge University Press

978-0-521-85935-6 - Exploitation of Fungi

Edited by G. D. Robson, P. van West and G. M. Gadd

Excerpt

[More information](#)

appressoria *ex planta*, and is highly amenable to genetic and molecular genetic manipulation (Dean, 1997; Talbot, 2003).

### Genome sequence

Efforts to sequence the genome began following the Second International Rice Blast Meeting held in Montpellier, France, in 1998. Here it was agreed that strain 70–15, a domesticated strain derived from Guy11, should be sequenced (Chao & Ellingboe, 1991). Initial efforts focused on creating deep BAC libraries, which were end sequenced, finger-printed and assembled into a physical map and integrated with the genetic map (Martin *et al.*, 2002; Zhu *et al.*, 1999). Other work resulted in the generation of >25 000 ESTs derived from a number of different growth conditions and stages (Ebbole *et al.*, 2004). More recently, a whole genome shotgun (WGS) approach was used to create a draft sequence. In all, greater than seven-fold sequence coverage was produced. The draft genome sequence consists of 2273 sequence contigs longer than 2 kb, ordered and oriented within 159 scaffolds. The total length of all sequence contigs is 38.8 Mb, and the total length of the scaffolds, including estimated sizes for the gaps, is 40.3 Mb (Dean *et al.*, 2005). Current efforts are focused on an optical map as an aid to genome assembly and closing the gaps in the genome sequence.

### Rapid expansion of repetitive elements

Novel pathogenic variants of *Magnaporthe* capable of infecting formerly resistant host plants arise with alarming frequency during rice cultivation. Such gains of virulence are often associated with transposon-mediated gene inactivation. Approximately 9.7% of the *Magnaporthe* genome assembly is comprised of repetitive DNA sequences. The majority of repetitive sequences in the assembly are retrotransposons comprising eight major families with at least 18 members. Five are retroelement families and three are DNA transposons. The repetitive elements are not uniformly distributed in the genome assembly, but form discrete clusters (Thon *et al.*, 2004).

These elements appear to have had a profound impact on genome evolution and also shed light on mechanisms of pathogenic variation. Investigation of the natural history of repetitive elements in *Magnaporthe* by the coalescent reconstructions revealed that most have undergone a rapid expansion in the genome at some point in their past. The ability to rapidly expand may be associated with the loss of repeat induced point (RIP) mutations, a genome defence mechanism that destroys duplicated sequences