Medical Mycology September 2006, **44**, S17–S21

Taylor & Francis
Taylor & Francis Group

Aspergillus niger genomics: Past, present and into the future

SCOTT E. BAKER
Fungal Biotechnology Team, Pacific Northwest National Laboratory, Richland, Washington, USA

Aspergillus niger is a filamentous ascomycete fungus that is ubiquitous in the environment and has been implicated in opportunistic infections of humans. In addition to its role as an opportunistic human pathogen, *A. niger* is economically important as a fermentation organism used for the production of citric acid. Industrial citric acid production by *A. niger* represents one of the most efficient, highest yield bioprocesses in use currently by industry. The genome size of *A. niger* is estimated to be between 35.5 and 38.5 megabases (Mb) divided among eight chromosomes/linkage groups that vary in size from 3.5–6.6 Mb. Currently, there are three independent *A. niger* genome projects, an indication of the economic importance of this organism. The rich amount of data resulting from these multiple *A. niger* genome sequences will be used for basic and applied research programs applicable to fermentation process development, morphology and pathogenicity.

Keywords *Aspergillus niger*, genome, wildtype, mutant

Introduction

Aspergillus niger is a filamentous ascomycete fungus that is ubiquitous in the environment and has been implicated in opportunistic infections of humans [1]. *A. niger* is most widely known for its role as a citric acid producer [2]. With production of citric acid at over one million metric tons annually, *A. niger* citric acid production serves as a model fungal fermentation process. As a common member of the microbial communities found in soils, *A. niger* plays a significant role in the global carbon cycle. This organism is a soil saprobe with a wide array of hydrolytic and oxidative enzymes involved in the breakdown of plant lignocellulose. A variety of these enzymes from *A. niger* are important in the biotechnology industry. *A. niger* is also an important model organism for several important research areas including the study of eukaryotic protein secretion in general, the effects of various environmental factors on suppressing or triggering the export of various biomass degrading enzymes, molecular mechanisms critical to fermentation process development, and mechanisms involved in the control of fungal morphology.

Currently, the genomes of three different strains of *A. niger* have been sequenced (Table 1). Two of the strains sequenced, NRRL 3 (ATCC 9029, CBS 120.49, N400) and ATCC 1015 (NRRL 328, CBS 113.46) are wildtype strains, while the other strain CBS 513.88, a derivative of NRRL 3122 (ATCC 22343, CBS 115989) was isolated after mutagenesis and selection for improved glucoamylase production [S.W. Peterson, personal communication]. Most recently, in 2005, the genome of *A. niger* ATCC 1015, a wildtype, historic strain was used in research that resulted in the first patented citric acid process [3], that was accepted for sequencing through the US Department of Energy (DOE) Microbial Genome Program (MGP). Organisms accepted by this program are sequenced by the DOE's Joint Genome Institute (JGI). Another wildtype *A. niger* strain, NRRL 3, was sequenced by Integrated Genomics, a US based company. Finally, CBS 513.88, a derivative of a mutant strain, NRRL 3122 was sequenced by a Netherlands based company, DSM [4,5].

Correspondence: Scott E. Baker, Fungal Biotechnology Team, Pacific Northwest National Laboratory, 902 Battelle Blvd., MSIN: K2-12, Richland, Washington 99352, USA. Tel: +1 509 372 4759. Fax: +1 509 372 4732. E-mail: scott.baker@pnl.gov

DOI: 10.1080/13693780600921037

Table 1 Comparison of genome projects.

Aspergillus niger genome sequencing program comparison Sponsors →Characteristics ↓	Integrated Genomics	DSM	DOE OBER Microbial Genome Program/ Joint Genome Institute (JGI)
Strain	Wild-type (NRRL 3)	Mutant (CBS 513.88)	Wild-type (ATCC 1015)
Project chronology	2000	2000 Public release, pending	2005 Draft sequence public released, April 2006; Ongoing sequence gap fillingfinishing
Aspergillus niger basepairs in scaffolds	33.7 Mb	35.9 Mb	37.2 Mb
Sequencing method	Shotgun	BAC tiling	Shotgun
Coverage	~6 ×	~7.5 ×	~8.9 ×
Genomic library insert size	1–2 kb	BAC	3 kb 8 kb 40 kb
Number of contigs or scaffolds	>9000 contigs	19 scaffolds	143 scaffolds
Estimated number of predicted genes	~14,000	~14,000	~11,200

Wildtype strains, ATCC 1015 and NRRL 3

In collaboration with the US Department of Energy (DOE) Joint Genome Institute (JGI), the Pacific Northwest National Laboratory Fungal Biotechnology Team has led a genome sequencing project for *A. niger*, strain ATCC 1015. Currently, an ~8.9X shotgun sequence assembly has been generated by the JGI. The ~8.9X coverage was generated from three different genomic libraries, with inserts of 3 kb, 8 kb and 40 kb. Paired end sequences from the 3 kb and 8 kb libraries each account for approximately 4X coverage, while the 40 kb library was sequenced to roughly 0.5 × coverage, also with paired ends. The resulting draft sequence assembled into less than 150 scaffolds. Both automated annotation and genome finishing (gap closure) efforts are underway. Manual gene annotation began in April 2006 after the release of the genome sequence (http://www.jgi.doe.gov/aspergillus with deposit into Genbank upon completion of manual annotation).

In parallel with the genome sequencing project, an *A. niger* (ATCC 1015) expressed sequence tag (EST) project was also launched at the JGI. Two sets of approximately 15,000 ESTs were sequenced from cDNA libraries generated from RNA pooled from an *A. niger* citric acid production time course (24, 48, 72, 96 and 120 h after spore inoculation) and RNA pooled from *A. niger* vegetative growth under a wide variety of nutritional conditions (corn fiber, rapeseed meal, a mix of purified lignin, cellulose, hemicellulose and protein, wheat bran, glucose, lactose, arabinose, starch/maltose, xylose, carbon limited, nitrogen/carbon limited). Combined with sequences already available in GenBank [6], the JGI generated ESTs will be used to aid in gene annotation.

ATCC 1015 is also notable as the parent strain of ATCC 11414 (NRRL 2270), an *A. niger* strain used in citric acid production studies [7,8]. ATCC 11414 was derived from a subculture of ATCC 1015 that displayed improved characteristics with regard to citric acid production [7].

The genome of another *A. niger* wildtype strain which produces high quantities of gluconic acid, NRRL 3 [9,10], was sequenced by Integrated Genomics, a company based in the United States. The sequence was purchased in 2004 by the Fungal Biotechnology Team at the Pacific Northwest National Laboratory under a Laboratory Directed Research and Development (LDRD) program whose research program centered on biobased products from fungal bioprocesses. Interestingly, the *A. niger* strain N402 (ATCC 64974) which has been used for mitotic recombination studies and generation of a genetic map was derived as a 'short conidiophore' mutant from *A. niger* strain NRRL 3 [11,12]. The approximate 6 × shotgun sequence coverage of *A. niger* strain NRRL 3 genome was generated from paired end sequences from a 1–2 kb insert genomic library. The assembly of the sequence was accomplished using the PHRAP assembly package and resulted in over 9000 contigs (paired end information was not used in the assembly).

Protein production strain, *A. niger* strain CBS 513.88

A Dutch company, DSM, sequenced the genome of the *A. niger* strain, CBS 513.88 [4,5]. This strain was

derived from the high glucoamylase producing strain, NRRL 3122 [4], itself an 'induced mutant'. Grown on agar, it has reduced conidiation and conidiophore morphology when compared to the wildtype *A. niger* strains ATCC 1015 and NRRL 3 (Baker, unpublished observations). Unlike the whole genome shotgun sequencing approach used by Integrated Genomics and the JGI, DSM used a BAC tiling approach resulting in a high quality draft sequence [5]. DSM announced plans for public release of the CBS 513.88 genome sequence at the 23rd Fungal Genetics Conference at Asilomar in 2005.

A. niger strain NRRL 3122 was deposited in 1964 by a researcher from the Fermentation Research Institute in Ministry of International Trade and Industry located in Chiba, Japan [S.W. Peterson, personal communication]. The history of the strain prior to the deposit is not readily available. Thus, it cannot be formally ruled out that NRRL 3122 was derived from either NRRL 3 or ATCC 1015.

The JGI *A. niger* ATCC 1015 genome sequence

The DOE JGI has released an 8.9 × shotgun draft sequence of the *A. niger* ATCC 1015 genome sequence. The DOE JGI *A. niger* strain ATCC 1015 draft genome sequence is available from www.jgi.doe.gov/aspergillus. Upon completion of sequence 'finishing' and manual annotation, gene models and the genome sequence will be deposited into GenBank. In its current assembly, the 8 chromosome genome is divided up into 143 scaffolds. The predicted size of the genome is approximately 37.1 Mb with approximately 11,200 predicted genes. The genome size is consistent with the 35.5–38.5 Mb predicted by electrophoretic karyotyping studies [13–15].

One of the most intriguing findings from the genome is the presence of a fumonisin-like biosynthetic gene cluster in the *A. niger* genome (Table 2). Fumonisin, an economically important secondary metabolite most notably produced by *Fusarium verticillioides* is a sphinganine analog mycotoxin (SAM) that inhibits the action of the enzyme ceramide synthase. AAL toxin, produced by *Alternaria alternat*a is another characterized SAM that is a virulence factor for infection of tomatoes. A search of the sequences generated by the *A. niger* (strain N400) EST sequencing project at Concordia University indicates evidence that genes in this cluster are expressed under one or more the growth conditions used in generation of the cDNA [6].

Conclusions and future directions

With a total three genome sequence databases soon to be publicly available, there is great potential to learn a lot about the biology of *A. niger* in the fields of protein secretion, fermentation and process development and opportunistic infections of humans. The very fact that there are three genome sequences from at least two

Table 2 A fumonisin-like gene cluster in *Aspergillus niger*.

Scaffold 1 coordinates (http://www.jgi.doe.gov/aspergillus)	Putative *Fusarium verticillioides* fumonisin ortholog	Orientation	Predicted gene function
1974582–1981918	None	Plus	Polyketide synthase
1982299–1983981	None	Minus	MFS transporter
1985577–1993676	Fum1	Plus	Polyketide synthase
1993867–1998834	Fum19	Minus	ABC transporter
1999513–2001227	Fum15	Minus	Cytochrome P450
2002023–2002527	None	Minus	Unknown
2003189–2005115	None	Plus	Unknown
2006028–2008035	Fum14	Plus	Amino acid Condensation domain
2008377–2009427	Fum13	Minus	Nucleoside-diphosphate-sugar Epimerase
2009952–2012602	Fum8	Minus	Acyltransferase
2013271–2014210	Fum3	Minus	Phytanoyl-Coa Dioxygenase
2014514–2015849	Fum7	Minus	Iron-containing alcohol Dehydrogenase
2016226–2017987	Fum10	Plus	AMP-dependent Synthetase and ligase
2018223–2019780	None	Minus	3-Ketosphinganine Reductase
2019891–2023317	Fum6	Plus	Cytochrome P450
2023679–2024973	None	Minus	Unknown
2025407–2026605	None	Minus	Alkaline Phytoceramidase
2018748–2030170	None	Plus	Unknown

distinct strain lineages will enable identification of single nucleotide polymorphisms (SNPs). The SNPs that are identified could be used in population genetic studies of both *A. niger* environmental and patient isolates. Furthermore, SNPs could be useful in improving the *A. niger* genetic map for use in mapping the genetic differences between wildtype and improved strains. It will be especially interesting to compare the sequences of the *A. niger* wildtype strains with the CBS 513.88 protein production strain that has undergone mutagenesis. While it may prove hard to separate natural variation from the effects of mutagens, this comparison could generate interesting data regarding the genes that are important for protein secretion.

The release of each *A. niger* genomes sequence database will accelerate molecular genetic analysis of genes involved in morphology and cell signaling, two processes important for *A. niger* as both a pathogen and fermentation organism. Molecular genetic techniques for analysis of genes in *A. niger* are well established; the sequenced *A. niger* genome will allow researchers the ability to quickly identify and delete, tag and/or overexpress genes and gene families whose role in human pathogenicity has been established in other aspergillosis causative species such as *Aspergillus fumigatus*.

The genus *Aspergillus* is well represented with regard to completed and in-progress genome sequencing projects. Recently, analyses of three Aspergillus genomes, *A. fumigatus*, *Emericella nidulans* and *Aspergillus oryzae* were published [16–18]. Genome projects for *Aspergillus flavus*, *Aspergillus clavatus*, *Aspergillus terreus* and *Neosartorya fischeri* are ongoing. The plethora of Aspergilli genome sequencing projects is a strong foundation on which to build comparative genomics studies. As more genome projects move from sequencing into data analysis, a high level snapshot of each organism's biology will emerge. Elucidation of the major differences at the genome level between species whose potential to infect humans differs greatly has the potential to open up novel avenues of research that might not have been obvious previously. Already, we know that the size of Aspergilli genomes may vary by up to 10 Mb [16–18].

With the release of the JGI generated *A. niger* genome sequence database and the pending release (as of this writing) of the DSM generated *A. niger* genome sequence database, the community of researchers with interest in this organism are poised to make important contributions that span across basic biological research, evolutionary biology and industrial and medical mycology. The availability of genome sequence data and gene model prediction allow proteomic and transcriptomic analysis at a global level to become much more straight forward and more accessible. The important test that lies ahead of these researchers will be to successfully synthesize and fully utilize the different types of data – proteome, transcriptome, metabolome – that spin out of the *A. niger* genome sequence.

Acknowledgements

The author wishes to thank Linda Lasure for helpful advice and Jon Magnuson, Ken Bruno and Ellen Panisko for their excellent review of the manuscript. The author also wishes to thank the scientists at the DOE Joint Genome Institute where this project is ongoing. Sequencing of A. niger ATCC 1015 performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program and the by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC03-76SF00098 and Los Alamos National Laboratory under contract No. W-7405-ENG-36.

References

1 Perfect JR, Cox GM, Lee JY, *et al*. The impact of culture isolation of *Aspergillus* species: a hospital-based survey of aspergillosis. *Clin Infect Dis* 2001; **33**: 1824–1833.
2 Magnuson J, Lasure L. Organic acid production by filamentous fungi. In: Tkacz J, Lange L (eds). *Advances in Fungal Biotechnology for Industry, Agriculture, and Medicine*. New York: Kluwer Academic & Plenum Publishers, 2004: 307–340.
3 Currie JN. Citric acid fermentation. *J Biol Chem* 1917; **31**: 15–37.
4 van Dijck PWM, Selten GCM, Hempenius RA. On the safety of a new genration of DSM *Aspergillus niger* enzyme production strains. *Reg Tox Pharm* 2003; **38**: 27–35.
5 van de Vondervoort PJI, Langeveld SMJ, Ram AFJ, *et al*. Comparison of the *Aspergillus niger* genomic DNA sequence with its genetic map. In: *The Second Aspergillus Meeting*. CA: Asilomar Conference Center, 2005.
6 Semova N, Storms R, John T, *et al*. Generation, annotation, and analysis of an extensive *Aspergillus niger* EST collection. *BMC Microbiol* 2006; **6**: 7.
7 Perlman D, Kita DA, Peterson WA. Production of citric acid from cane molasses. *Arch Biochem* 1946; **11**: 123–129.
8 Dai Z, Mao X, Magnuson J, Lasure L. Identification of genes associated with morphology in *Aspergillus niger* by using suppression subtractive hybridization. *Appl Env Micro* 2004; **70**: 2474–2485.
9 Blom RH, Pfeifer VF, Moyer AJ, *et al*. Sodium gluconate production fermentation with *Aspergillus niger*. *Ind Eng Chem* 1952; **44**: 435–440.
10 Moyer AJ, Umberger EJ, Stubbs JJ. Fermentation of concentrated solutions of glucose to gluconic acid improved process. *Ind Eng Chem* 1940; **32**: 1379–1383.

11 Bos CJ, Debets AJ, Swart K, *et al*. Genetic analysis and the construction of master strains for assignment of genes to six linkage groups in *Aspergillus niger*. *Curr Genet* 1988; **14**: 437–443.
12 Debets F, Swart K, Hoekstra RF, Bos CJ. Genetic maps of eight linkage groups of *Aspergillus niger* based on mitotic mapping. *Curr Genet* 1993; **23**: 47–53.
13 Verdoes JC, Calil MR, Punt PJ, *et al*. The complete karyotype of *Aspergillus niger*: the use of introduced electrophoretic mobility variation of chromosomes for gene assignment studies. *Mol Gen Genet* 1994; **244**: 75–80.
14 Debets AJ, Holub EF, Swart K, van den Broek HW, Bos CJ. An electrophoretic karyotype of *Aspergillus niger*. *Mol Gen Genet* 1990; **224**: 264–268.
15 Swart K, Debets AJ, Bos CJ, *et al*. Genetic analysis in the asexual fungus *Aspergillus niger*. *Acta Biol Hung* 2001; **52**: 335–343.
16 Galagan JE, Calvo SE, Cuomo C, *et al*. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 2005; **438**(7071): 1105–1115.
17 Machida M, Asai K, Sano M, *et al*. Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* 2005; **438**(7071): 1157–1161.
18 Nierman WC, Pain A, Anderson MJ, *et al*. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* 2005; **438**(7071): 1151–1156.