### Mini-Review

Correspondence Meriel G. Jones m.g.jones@liv.ac.uk

## The first filamentous fungal genome sequences: Aspergillus leads the way for essential everyday resources or dusty museum specimens?

Meriel G. Jones

The University of Liverpool, School of Biological Sciences, Biosciences Building, Crown Street, Liverpool L69 7ZB, UK

The published Aspergillus genome sequences (*A. nidulans, A. fumigatus, A. oryzae*) and further sequence data from *A. clavatus, Neosartorya fischeri, A. flavus, A. niger, A. parasiticus* and *A. terreus* are the first from a group of related filamentous fungi. They indicate the gains possible from genomic approaches, but also problems that arise after the sequences are finished. Benefits include a greater understanding of genome structure and evolution, insights into gene regulation, predictions of new factors that may be relevant to pathogenicity and the discovery of novel enzymes with biotechnological value. Areas where further developments are needed include gene and structure–function predictions, methods for comparative genome analysis and the interfaces for access to genome information. In addition, strategies for continued maintenance and updating need to be developed at the start of the post-genomic era to increase the value of genome sequences into the future.

#### Introduction

There has been substantial investment in sequencing and initial annotation of the first filamentous fungal genomes. This collection will expand substantially in the next few years. Genomics gives new tools for fungal biology that should allow more rapid transfer of ideas, knowledge and expertise between specialist areas. Now is therefore a good time to consider how to make best use of these data as postgenomics becomes less glamorous and more open-ended.

The aspergilli provide a good example of what can be gained from genome sequences and highlights some of the problems. The interest in this group comes from their importance as pathogens and allergens of humans [Aspergillus fumigatus; Neosartorya fischeri (teleomorph of A. fischerianus), its little-studied, non-pathogenic close relative; A. flavus; and the very rarely pathogenic A. clavatus and A. terreus], plant pathogens (A. flavus; A. parasiticus), industrial organisms (A. oryzae; A. niger; A. terreus) and as a model organism (A. nidulans). So far, the genome sequence of nine species has been partially or fully completed. The sequences of three (A. nidulans, A. fumigatus, A. oryzae) were published at the end of 2005, with the sequence of one more (A. niger) completed but not publicly available (Galagan et al., 2005a; Nierman et al., 2005; Machida et al., 2005). Assembled genome sequences for A. clavatus, A. flavus, A. terreus, N. fischeri and further strains of A. fumigatus and A. niger are also at or near completion and there are also some data for A. parasiticus (see Table 1; 'Aspergillus Genomics' on the Aspergillus website http:// www.aspergillus.org.uk/indexhome.htm; 2006; Baker,

Payne *et al.*, 2006; Wortman *et al.*, 2006). The initial genomic and first post-genomic analyses bring out the following issues.

# Fungal genome data and the research community

Convenient worldwide access to the genome data is an obvious requirement, with a need for continued maintenance and additions. As an example, the current A. nidulans genome website focuses on the outcome of the sequencing project, providing access to gene predictions and initial annotation, views of genome features, downloads of DNA sequences and predicted protein sequences and a BLAST search facility (http://www.broad.mit.edu/annotation/fungi/aspergillus/index.html). Other web resources for A. nidulans, such as the gene lists and linkage maps curated at the University of Glasgow, UK (http://www.gla. ac.uk/Acad/IBLS/molgen/aspergillus/), and the genetic resources maintained at the Fungal Genetics Stock Center, Missouri, USA (http://www.fgsc.net/asperg.html), remain to be integrated. The benefits of continued updating can be seen from the Saccharomyces cerevisiae genome site (http:// www.yeastgenome.org), where sequencing was ostensibly completed a decade ago. Nevertheless, the database is still being corrected for sequence, splicing and other fundamental features as well as providing increasingly detailed functional annotation. The integration of additional datasets, including microarray, proteomic and metabolomic studies, adds value to the genome sequence. One additional aspect that needs consideration is the way in which the data

#### Table 1. Websites and status of the Aspergillus genomes that have been or are being sequenced (November 2006)

For a fuller list of all public fungal genome sequencing projects, see Galagan *et al.*, 2005b, and the NCBI-ENTREZ genome projects web pages (http://www.ncbi.nlm.nih.gov/entrez/ query.fcgi?db=genomeprj).

Species	Strain	URL	Status of genome sequence and collaborators (November 2006)	Reference
A. clavatus	NRRL 1 Type strain	http://msc.tigr.org/aspergillus/aspergillus_clavatus_nrrl_1/index.shtml	Assembled. Sequencing and annotation at TIGR. Collaboration with Broad.	Wortman et al. (2006)
A. flavus	NRRL 3357	http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species = a_flavus; http://www.aspergillusflavus.org/	Assembly. TIGR/University of Oklahoma. USDA/ARS.	Payne et al. (2006)
A. fumigatus	Af293	http://www.tigr.org/tdb/e2k1/afu1/	Annotated. TIGR/Sanger/Institut Pasteur.	Nierman et al. (2005)
		http://www.cadre.man.ac.uk/Aspergillus_fumigatus/	Curated genome.	
	CEA10		Merck & Co.	Wortman et al. (2006)
A. nidulans	FGSC A4	http://www.broad.mit.edu/annotation/genome/aspergillus_nidulans/Home.html	Annotated. Broad Institute of MIT and Harvard (http://www.broad.mit.edu), Monsanto.	Galagan <i>et al.</i> (2005a)
		http://www.cadre.man.ac.uk/Aspergillus_nidulans/	Curated genome.	
A. niger	ATCC 1015	http://genome.jgi-psf.org/Aspni1/Aspni1.home.html	Annotated. DOE Joint Genome Institute with 14 collaborators.	Baker (2006)
	NRRL 3		Integrated Genomics.	Baker (2006)
	CBS 513.88	http://www.dsm.com/en_US/html/dfs/genomics_aniger.htm	DSM licenced access, public release pending.	
A. oryzae	RIB40	http://www.bio.nite.go.jp/dogan/MicroTop?GENOME_ID = ao	Annotated. Department of Biotechnology, National Institute of Technology and Evaluation, Japan.	Machida <i>et al.</i> (2005); Payne <i>et al.</i> (2006)
A. terreus	NIH 2624	http://www.broad.mit.edu/annotation/genome/aspergillus_terreus/Home.html	Annotated. Broad.	Wortman et al. (2006)
	ATCC 20542	http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val = AABT00000000	Assembly. Microbia Inc.	Askenazi et al. (2003)
A. parasiticus		http://www.genome.ou.edu/fungal.html	In progress. University of Oklahoma.	
Neosartorya fischerii (teleomorph of A. fischerianus)	NRRL 181	http://msc.tigr.org/aspergillus/neosartorya_fischeri_nrrl_181/index.shtml	Assembled. Sequencing and annotation at TIGR. Collaboration with Broad.	Wortman et al. (2006)

are organized. Convergence towards one web-interface style for displaying and manipulating genomic data would lower the learning threshold for researchers beginning to take a genomic view of their favourite filamentous fungus.

It is unrealistic to expect that the same level of resource will be available for every fungus. Of the 1.6 million fungal species that may exist, only about 80 000 have been described at some scientific level (Hawksworth, 1991). The number studied in detail is even smaller, with researchers frequently focusing on only one particular feature of a species. The list of filamentous fungi sequenced to date comprises many of the well-known major model, pathogenic and economically important species. In addition to aspergilli they include Botrytis cinerea, Chaetomium globosum, Coprinus cinereus, Fusarium graminearum, Gibberella zeae, Magnaporthe grisea, Neurospora crassa, Phanerochaete chrysosporium, Podospora anserina, Rhizopus oryzae, Sclerotinia sclerotiorum, Trichoderma reesei and Ustilago maydis (see listing and current status of public fungal genome projects on NCBI-ENTREZ: http:// www.ncbi.nlm.nih.gov/entrez/query.fcgi?db = genomeprj; Fungal Genome Initiative: http://www.broad.mit.edu/ annotation/fgi/; DOE Joint Genome Institute: http:// www.jgi.doe.gov/; Podospora anserina Genome Project: http://podospora.igmors.u-psud.fr/index.html; Galagan et al., 2005b). Sequencing of further important species is under way, including several mycorrhizal species (see for example International Mycorrhiza Genome Consortium: http://mycor.nancy.inra.fr/ectomycorrhizadb/; Basidiomycete Genome projects: http://www.basidiomycetes.org/geno.htm). However, most of the fungal genomes that are now being sequenced, and which will be sequenced in future, rely on very small resources of human expertise and existing information.

The difficulties experienced in funding living microbial culture collections and curating museum specimens, despite their value for scientific research, industry and human culture, indicates a path to avoid. A solution may come from the electronic nature of genomic data so that strategies for many members of the research community to maintain and develop genome resources over decades are feasible. The start of the genomic era for filamentous fungi is the time to implement such strategies. The Eurofungbase project (http://www.eurofung.net/), supported by the European Union, is an example of one initiative within the Aspergillus research community. It aims to use the Aspergillus genome sequences to draw together the fragmented knowledge on aspergilli to benefit researchers working with each species within the fields of biotechnology and health. Events such as annotation jamborees are planned to harness expertise within the community.

#### Genome annotation

An accurate catalogue of the genes within the genome is an obvious goal that would greatly facilitate post-genomic research, and the Aspergillus genomes fall short of this at the moment. The initial genome analysis of A. nidulans predicted that there were 9541 protein-coding genes, which was revised to 10701 in release 4 (March 2006) after extensive manual revision. Methods for accurate gene prediction are therefore one fundamental requirement. Some classes of genes present in filamentous fungi are almost entirely absent from other organisms. These include developmental pathways involving extensive protein secretion, the hyphal growth form and fungal reproductive structures (Fungal Genome Initiative Steering Committee, 2003). Now that several genomes have been finished, strategies for predictions in further filamentous fungi should become easier. The generation and integration of substantial EST data during each sequencing project would help gene prediction significantly, as well as giving insights into alternative splicing in each species (Galagan et al., 2005b).

Predicting function from sequence continues to be a challenge, especially within secondary metabolism, which is the reason for sustained academic, commercial and medical interest in several aspergilli. The focused research within secondary metabolites, such as statin biosynthesis in *A. terreus*, provides expertise and resources with a reason to address specific pathways and classes of genes. The role of toxins in *A. fumigatus* pathogenicity is another long-standing research question and the genome sequence has already aided location of biosynthetic genes for some of these toxins (Maiya *et al.*, 2006).

#### Comparative genome analysis

Comparative analysis will be an essential part of the interpretation of future fungal genomes, especially since the specific interests of many small research communities mean that organism-specific data are limited. Initial comparisons between the first three Aspergillus sequences have shown the value of this approach, as well as indicating the need for improved tools (Galagan et al., 2005a). Several further Aspergillus species were selected for sequencing to enable detailed comparisons between genomes of much more closely related species that differed in pathogenicity or other properties (Fungal Genome Initiative Steering Committee, 2003). The preliminary comparison of A. fumigatus with A. clavatus and N. fischeri suggests that these hopes have been fulfilled (Wortman et al., 2006). The genomes of A. oryzae and A. flavus also appear very similar, with both enriched for secondary metabolism, although the metabolome of A. flavus appears more complex (Payne et al., 2006). Comparative analysis of A. nidulans, A. fumigatus and A. oryzae led to the realization that it may be possible to develop the power of genetic analysis in the latter species because the genes involved in the mating process in A. nidulans are also found in the others (Galagan et al., 2005a; Nierman et al., 2005).

Although secondary metabolites are the reason for interest in many fungi, if these gene clusters lie within heavily rearranged subtelomeric regions, as in the first sequenced aspergilli, comparative methods may not be helpful (Galagan *et al.*, 2005a). Interestingly, the variation between two sequenced strains of *A. fumigatus* was located primarily in subtelomeric regions, indicating that this area may be important in maintaining species variability (Wortman *et al.*, 2006). Nevertheless, identification of the putative cluster of genes necessary for the biosynthesis of gliotoxin in *A. fumigatus* was possible through a comparative genomics approach (Gardiner & Howlett, 2005).

The development of infrastructure for comparative functional genomics in fungal species, such as through the eFungi project (http://www.e-fungi.org.uk/) funded by a UK Research Council, is therefore a welcome step. This project aims to facilitate systematic comparisons of fungal genomes through developing a data warehouse to integrate data from several species, initially 23 fungi (including several yeast and dimorphic species) and two oomycetes. A library of bioinformatics queries is being developed to analyse the data, so that less-well-studied species can be readily compared with better-understood species. The first studies from the project will test the effectiveness of this comparative genomic infrastructure and are planned to focus on the scientific question of the nature and origin of fungal pathogenicity.

Comparative analysis will also lead to deeper understanding of the control of gene expression, building on knowledge from a few well-studied systems (e.g. Caddick, 2004). The significance of conserved non-coding regions, as well as sets of co-regulated genes with no apparent conserved upstream or downstream motifs, is where combined genome analyses and experimental studies will yield insights that are not apparent from studies of individual genes or species (e.g. Galagan *et al.*, 2005a).

#### Tools for functional analysis

The small size and specialized nature of the research communities working with most filamentous fungi provides only a fragmentary picture of their biology. The model fungi, such as A. nidulans, give a more rounded view as well as insights into fundamental areas of all eukaryotic biology. Research with A. nidulans started in the 1950s (Pontecorvo et al., 1953) and continues to advance understanding of metabolic regulation, pH control, DNA repair, controlled RNA degradation, morphogenesis, chromatin structure, cell cycle control, cytoskeletal function, mitochondrial DNA structure and other aspects of eukaryotic genetics and cell physiology. The existence of a well-characterized sexual cycle (teleomorph Emericella nidulans) and molecular biology tools make it a valuable laboratory species. A well-annotated and supported genome sequence will take A. nidulans into the post-genomic era, repaying the effort that will be needed to achieve this.

The availability of tools for more rapid functional analysis is important but many rely on a well-annotated genome. There have been technical developments in tagging and

deletion by homologous integration in Aspergillus species so that the development of large-scale collections of mutants is becoming feasible (e.g Yang et al., 2004; Ferreira et al., 2006; Krappmann, 2006; Navak et al., 2006). Several types of microarrays have been developed and used for the first global views of gene expression in aspergilli. A. nidulans microarrays have added to the already complex picture of carbon catabolite repression as mediated by creA (Mogensen et al., 2006) while, in combination with manipulation of expression of laeA, a global regulator of secondary metabolism, microarrays have started to reveal the secondary metabolome (Bok et al., 2006). The technology for proteomics is still developing, but the first applications to A. fumigatus have shown that different carbon sources cause changes to the proteome (Kniemeyer et al., 2006).

Before the number of these studies increases, the question of how to get more value from the raw data through making it accessible for reanalysis by future researchers needs to be answered. Data standards are emerging but a location and responsibility for maintenance of these post-genomic resources needs thought.

#### Applied versus pure science

The life sciences and biotechnology are widely recognized as the next stage in the knowledge-based economy, creating new work, social and entrepreneurial opportunities. The status of aspergilli as pathogens and industrial organisms means that their genomes are important for commercial as well as academic purposes and several firms have started sequencing projects (see Table 1). Indeed, the first completed sequencing and annotation project for *A. niger* was funded privately by the major Dutch food and pharmaceutical ingredients company DSM and provides an example of the real commercial potential of fungal genome sequences. Commercial exploitation of the *A. oryzae* genome is also well under way (Abe *et al.*, 2006).

A. niger is important as the major world source of citric acid and higher-value enzyme products including pectinases, proteases, amyloglucosidases, cellulases, hemicellulases and lipases. Enzymes from A. niger, including ones from recombinant strains where the fungus is used for expression, have been repeatedly assessed as 'generally regarded as safe' by the World Health Organization and US Food and Drug Administration (Schuster et al., 2002). DSM's 2005 annual company report (Royal DSM N.V., 2005) indicated that genomic information has been used to optimize production of 7-aminodeacetoxycephalosporanic acid, a commercial intermediate for semisynthetic penicillins and cephalosporins. Over 200 new proteases were identified during annotation and at least two have now reached commercial products. These are 'Brewer's Clarex', a proline-specific endoprotease processing aid to prevent chill-haze in beer (Lopez & Edens, 2005), and a protease to produce milk peptides for new 'Peptopro' and 'Multipower' sports drinks (Edens et al., 2005). One consequence of limited access to

A. *niger* genome information has been the initiation of several public sequencing projects for this species (see Table 1 and Baker, 2006). The resulting complete and partial sequences unintentionally provide data on intraspecific variation.

Exploitation of the genomes of *A. fumigatus*, *N. fischeri* and others could lead to new medical products as well as advancing understanding of the nature of Aspergillus pathogenicity (Krappmann, 2006; Wortman *et al.*, 2006). The world market for antifungal drugs is estimated to reach £4.6 billion in 2007, even though many have serious side-effects and are ineffective against species such as *A. fumigatus*, and antifungal resistance is increasing. There is also a lack of effective rapid diagnostic tests. Post-genomics offers new approaches to these problems.

Reconciling differing academic and commercial objectives will continue to be one aspect of the exploitation of fungal genomes. This is immediately obvious in species that are already of commercial value, but advances in comparative genomics and functional prediction may give new monetary value to currently worthless facts.

#### The future challenge

We are at the start of post-genomic biology in the filamentous fungi. The first genome sequences of the aspergilli have provided insights into their biology and leads for new commercial products. However, they have also highlighted the problems that arise after the sequencing is over, especially if fungal genomes are to become everyday tools for all biologists rather than the province of specialist genome scientists. After succeeding with the sequencing challenge, the scientific community has to resolve how to retain value into the future.

#### Acknowledgements

I am grateful to the Eurofungbase project LSSG-CT-2005-018964 and the anonymous referees for advice.

#### References

Abe, K., Gomi, K., Hasegawa, F. & Machida, M. (2006). Impact of *Aspergillus oryzae* genomics on industrial production of metabolites. *Mycopathologia* **162**, 143–153.

Askenazi, M., Driggers, E. M., Holtzman, D. A., Norman, T. C., Iverson, S., Zimmer, D. P., Boers, M.-E., Blomquist, P. R., Martinez, E. J. & other authors (2003). Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. *Nat Biotechnol* 21, 150–156.

Baker, S. C. (2006). Aspergillus niger genomics: past, present and into the future. Med Mycol 44, S17–S21.

Bok, J. W., Hoffmeister, D., Maggio-Hall, L. A., Murillo, R., Glasner, J. D. & Keller, N. P. (2006). Genomic mining for Aspergillus natural products. *Chem Biol* 13, 31–37.

Caddick, M. X. (2004). Nitrogen regulation in mycelial fungi. In *The Mycota III, Biochemistry and Molecular Biology,* 2nd edn,

pp. 349–368. Edited by R. Brambl & G. A. Marzluf. Berlin & Heidelberg: Springer.

Edens, L., Dekker, P., Van der Hoeven, R., Deen, F., De Roos, A. & Floris, R. (2005). Extracellular prolyl endoprotease from *Aspergillus niger* and its use in the debittering of protein hydrolysates. *J Agric Food Chem* 53, 7950–7957.

Ferreira, M. E. D., Kress, M. R. V. Z., Savoldi, M., Goldman, M. H. S., Hartl, A., Heinekamp, T., Brakhage, A. A. & Goldman, G. H. (2006). The *akuB*(KU80) mutant deficient for nonhomologous end joining is a powerful tool for analyzing pathogenicity in *Aspergillus fumigatus*. *Eukaryot Cell* 5, 207–211.

**Fungal Genome Initiative Steering Committee (2003).** A whitepaper for comparative fungal genomics. www.broad.mit.edu/ annotation/fgi.

Galagan, J. E., Calvo, S. E., Cuomo, C., Ma, L. J., Wortman, J., Batzoglou, S., Lee, S. I., Basturkmen, M., Spevak, C. & other authors (2005a). Sequence of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 438, 1105–1115.

Galagan, J. E., Henn, M. R., Ma, L.-J., Cuomo, C. A. & Birren, B. (2005b). Genomics of the fungal kingdom: insights into eukaryotic biology. *Genome Res* 15, 1620–1631.

Gardiner, D. M. & Howlett, B. J. (2005). Bioinformatic and expression analysis of the putative gliotoxin biosynthetic gene cluster of *Aspergillus fumigatus. FEMS Microbiol Lett* 248, 241–248.

Hawksworth, D. L. (1991). The fungal dimension of biodiversity: magnitude, significance and conservation. *Mycol Res* 95, 641–655.

Kniemeyer, O., Lessing, F., Scheibner, O., Hertweck, C. & Brakhage, A. A. (2006). Optimisation of a 2-D gel electrophoresis protocol for the human-pathogenic fungus *Aspergillus fumigatus*. *Curr Genet* **49**, 178–189.

Krappmann, S. (2006). Tools to study molecular mechanisms of Aspergillus pathogenicity. *Trends Microbiol* 14, 356–364.

Lopez, M. & Edens, L. (2005). Effective prevention of chill-haze in beer using an acid proline-specific endoprotease from *Aspergillus niger*. J Agric Food Chem 53, 7944–7949.

Machida, M., Asai, K., Sano, M., Tanaka, T., Kumagai, T., Terai, G., Kusmoto, K., Arima, T., Akita, O. & other authors (2005). Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* **438**, 1157–1161.

Maiya, S., Grundmann, A., Li, S. M. & Turner, G. (2006). The fumitremorgin gene cluster of *Aspergillus fumigatus*: identification of a gene encoding brevianamide F synthetase. *Chem Biochem* 7, 1062–1069.

Mogensen, J., Nielsen, H. B., Hofmann, G. & Nielsen, J. (2006). Transcription analysis using high-density micro-arrays of *Aspergillus nidulans* wild-type and *creA* mutant during growth on glucose or ethanol. *Fungal Genet Biol* **43**, 593–603.

Nayak, T., Szewczyk, E., Oakley, C. E., Osmani, A., Ukil, L., Murray, S. L., Hynes, M. J., Osmani, S. A. & Oakley, B. R. (2006). A versatile and efficient gene-targeting system for *Aspergillus nidulans. Genetics* 172, 1557–1566.

Nierman, W. C., Pain, A., Anderson, M. J., Wortman, J. R., Kim, H. S., Arroyo, J., Berriman, M., Abe, K., Archer, D. B. & other authors (2005). Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* **438**, 1151–1156.

Payne, G. A., Nierman, W. C., Wortman, J. R., Pritchard, B. L., Brown, D., Dean, R. A., Bhatnagar, D., Cleveland, T. E., Machida, M. & Yu, J. (2006). Whole genome comparison of *Aspergillus flavus* and *A. oryzae. Med Mycol* 44, S9–S11.

Pontecorvo, G., Roper, J. A., Hemmons, L. M., Macdonald, K. D. & Bufton, A. W. (1953). The genetics of *Aspergillus nidulans*. *Adv Genet* 5, 141–238.

Royal DSM N. V. (2005). Annual Report.

Schuster, E., Dunn-Coleman, N., Frisvad, J. & van Dijck, P. (2002). On the safety of Aspergillus niger – a review. Appl Microbiol Biotechnol 59, 426–435.

Wortman, J. R., Fedorova, N., Crabtree, J., Joardar, V., Maiti, R., Haas, B. J., Amedeo, P., Lee, E., Angiuoli, S. V. & other authors

(2006). Whole genome comparison of the *A. fumigatus* family. *Med Mycol* 44, S3–37.

Yang, L., Ukil, L., Osmani, A., Nahm, F., Davies, J., De Souza, C. P. C., Dou, X. W., Perez-Balaguer, A. & Osmani, S. A. (2004). Rapid production of gene replacement constructs and generation of a green fluorescent protein-tagged centromeric marker in *Aspergillus nidulans. Eukaryot Cell* 3, 1359–1362.