The 2008 Update of the *Aspergillus nidulans* Genome Annotation: A Community Effort

Jennifer Russo Wortman  
*Department of Medicine Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA*

Steven D. Harris  
*University of Nebraska-Lincoln, Steven.Harris@umanitoba.ca*

Geoffrey Turner  
*University of Sheffield, Sheffield, UK, g.turner@sheffield.ac.uk*

Aspergillus nidulans consortium

Follow this and additional works at: [https://digitalcommons.unl.edu/plantscifacpub](https://digitalcommons.unl.edu/plantscifacpub)  
Part of the Plant Sciences Commons

Russo Wortman, Jennifer; Harris, Steven D.; Turner, Geoffrey; and consortium, Aspergillus nidulans, "The 2008 Update of the *Aspergillus nidulans* Genome Annotation: A Community Effort" (2009). *Faculty Publications from the Center for Plant Science Innovation*. 33.  
[https://digitalcommons.unl.edu/plantscifacpub/33](https://digitalcommons.unl.edu/plantscifacpub/33)

This Article is brought to you for free and open access by the Plant Science Innovation, Center for at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications from the Center for Plant Science Innovation by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
The 2008 update of the *Aspergillus nidulans* genome annotation: A community effort


*Corresponding author.
E-mail address: g.turner@sheffield.ac.uk (G. Turner).
Molecular Cell Biology, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Haren, The Netherlands
Section of Plant Biology, University of California, Davis, CA, USA
Vienna University of Technology, Institute of Chemical Engineering, Division of Gene Technology and Applied Biochemistry, Vienna, Austria
University of Kentucky, School of Biological Sciences, Lexington, KY, USA
Department of Microbial Biotechnology and Cell Biology, Faculty of Science and Technology, University of Debrecen, Debrecen, Hungary
Department of Molecular Sciences, University of Tennessee, Memphis, USA
Department of Chemical and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden
Department of Molecular Genetics, Ohio State University, Columbus, OH, USA
VTT Technical Research Centre of Finland, Espoo, Finland
Institute of Biochemistry and Biophysics, Department of Genetics, Warsaw, Poland
Department of Chemistry and Biomolecular Sciences, Macquarie University Sydney, Australia
TNO Quality of Life, Industrial Biotechnology, Zeist, The Netherlands
Institut de Génétique et Microbiologie, CNRS - Université Paris-Sud XI, Orsay, France
Faculty of Life Sciences, University of Manchester, Manchester, UK
Danisco-Genencor, Leiden, The Netherlands
Sandoz GmbH, Kundl, Tirol, Austria
Helmholtz Centre for Infection Research, Braunschweig, Germany
University of Karlsruhe, Department of Applied Biosciences Applied Microbiology, Karlsruhe, Germany
Center for Biological Sequence Analysis BioCentrum-DTU, The Technical University of Denmark, Lyngby, Denmark
Dyadic Nederland BV, Wageningen, The Netherlands
DSM Anti-Infectives, Delft, The Netherlands
Microbiology, Utrecht University, Utrecht, The Netherlands
DOE Plant Research Lab, Michigan State University, East Lansing, MI, USA
Department of Biochemistry, USUHS, Bethesda, MD, USA
Hamburg University of Technology, Institute of Bioprocess and Biosystems Engineering; Hamburg, Germany and Helmholtz Centre for Infection Research, Braunschweig, Germany
Centre for Integrative Bioinformatics (IBIVU), VU University of Amsterdam, The Netherlands
School of Computer Science, University of Manchester, Manchester, UK
Fungal Genetics and Technology consultancy, Wageningen, The Netherlands
Department of Biochemistry, University of Cambridge, Cambridge, UK
Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, UK

ARTICLE INFO

Article history:
Received 17 October 2008
Accepted 15 December 2008
Available online 25 December 2008

Index descriptors:
Aspergillus nidulans
Aspergilli
Genome
Annotation
Fungal community
Assembly
Transcription factors
CADRE

ABSTRACT

The identification and annotation of protein-coding genes is one of the primary goals of whole-genome sequencing projects, and the accuracy of predicting the primary protein products of gene expression is vital to the interpretation of the available data and the design of downstream functional applications. Nevertheless, the comprehensive annotation of eukaryotic genomes remains a considerable challenge. Many genomes submitted to public databases, including those of major model organisms, contain significant numbers of wrong and incomplete gene predictions. We present a community-based reannotation of the Aspergillus nidulans genome with the primary goal of increasing the number and quality of protein functional assignments through the careful review of experts in the field of fungal biology. © 2009 Elsevier Inc. All rights reserved.

1. Introduction

1.1. Importance of Aspergillus nidulans as a reference for the eight sequenced aspergilli

Genome sequences of eight species of the genus Aspergillus have now been determined (Table 1). However, the aspergillus research community is relatively small, and the task of detailed annotation of all of these genomes, together with presentation of the information in an easily accessible form, is enormous. We have therefore chosen to begin the task with the genetic model species A. nidulans.

1.1.1. General importance of aspergilli for fungal biology, human health, biotechnology, and agriculture

The genus Aspergillus comprises a diverse group of filamentous fungi. Despite belonging to the same genus, Aspergillus species have diverged significantly (Galagan et al., 2005), though they are sufficiently related such that orthologues can be identified for the majority of genes. A. nidulans, the most studied species, has been an important model organism for eukaryotic genetics for over 60 years (Martinelli and Kinghorn, 1994), with the advantage of having a sexual cycle (teleomorph Emericella nidulans), which is usually absent from other species of Aspergillus. In addition to a long history of classical genetic and biochemical studies, most molecular techniques for gene manipulation were first developed in A. nidulans before application to other members of the genus. Detailed laboratory protocols have recently been made easily accessible (Osmani et al., 2006; Szewczyk et al., 2006; Todd et al., 2007a; Todd et al., 2007b), and mutant strains isolated over many years are available from the Fungal Genetics Stock Centre (www.fgsc.net), together with other useful resources such as vectors and libraries.

It is important to note that different wild-type strains of A. nidulans exist (Jinks et al., 1966), but all the commonly used mutant strains are derived from a single strain, sometimes called the Glasgow strain, following its choice as a genetic model in the early 1950s (Pontecorvo et al., 1953). Our detailed understanding of the genetics and physiology of A. nidulans provides an excellent basis for extension of this knowledge to other, imperfect species of economic importance. These include the opportunistic pathogen Aspergillus fumigatus, a cause of allergies and a growing threat to immunocompromised patients; Aspergillus niger and Aspergillus oryzae, sources of industrial enzymes and other products such as citric acid; Aspergillus flavus, a plant pathogen and toxin-producing...
<table>
<thead>
<tr>
<th>Aspergillus species</th>
<th>Sequencers</th>
<th>Annotation methods</th>
<th>References</th>
<th>Links and accession numbers to sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. nidulans (Emericella nidulans FGSC A4)</td>
<td>Broad Institute</td>
<td>Calhoun annotation system, including FGENESH (Salamov and Solovyev, 2000); FGENESH+; and GENEWISE (Birney et al., 2004)</td>
<td>Galagan et al. (2005)</td>
<td>Genbank: AACD00000000 <a href="http://www.broad.mit.edu/annotation/genome/aspergillus_group/">http://www.broad.mit.edu/annotation/genome/aspergillus_group/</a></td>
</tr>
<tr>
<td>A. fumigatus (AI293)</td>
<td>The Institute for Genome Research (TIGR)</td>
<td>TIGR annotation pipeline, including GlimmerM; Exonomy; Unveil (Majeros et al., 2003); and GeneSplicer (Pertea et al., 2001) plus manual annotation</td>
<td>Nierman et al. (2005)</td>
<td>Genbank: AAHF00000000 <a href="http://www.tigr.org/tdb/e2k1/afu1/">http://www.tigr.org/tdb/e2k1/afu1/</a></td>
</tr>
<tr>
<td>A. niger (CBS 513.88)</td>
<td>DSM Food Specialties/ Biomax Informatics AG</td>
<td>Pedant-ProTM Sequence Analysis Suite plus manual annotation by Aspergillus research community</td>
<td>Pel et al. (2007)</td>
<td></td>
</tr>
<tr>
<td>A. niger (ATCC 1015)</td>
<td>Joint Genome Institute (JGI)</td>
<td>JGI annotation pipeline including FGENESH; FGENESH+; and GENEWISE JCVI annotation pipeline including GlimmerHMM (Majeros et al., 2003); Genezilla (Majeros et al., 2003); SNAP (Korf, 2004); Genewise; Twinscan (Flicek et al., 2003); PASA (Haa et al., 2003); and EvidenceModeler (Haas et al., 2008)</td>
<td>Fedorova et al. (2008)</td>
<td>GenBank: ABDI00000000 <a href="http://genome.jgi-psf.org/Aspni1/">http://genome.jgi-psf.org/Aspni1/</a></td>
</tr>
<tr>
<td>A. clavatus (NRRL1)</td>
<td>Joint Craig Venter Institute (JCVI)</td>
<td>JCVI annotation pipeline including GlimmerHMM; Genezilla; SNAP; Genewise; Twinscan; PASA; and EvidenceModeler</td>
<td>Fedorova et al. (2008)</td>
<td>GenBank: AAKD00000000 <a href="http://www.tigr.org/tdb/e2k1/acl1/">http://www.tigr.org/tdb/e2k1/acl1/</a></td>
</tr>
<tr>
<td>A. fischeriana (Neosartorya fischeri NRRL181)</td>
<td>Joint Craig Venter Institute (JCVI)</td>
<td>JCVI annotation pipeline including GlimmerHMM; Genezilla; SNAP; Genewise; Twinscan; PASA; and EvidenceModeler</td>
<td>Fedorova et al. (2008)</td>
<td>GenBank: AAJN00000000 <a href="http://www.aspergillusflavus.org/genomics/">http://www.aspergillusflavus.org/genomics/</a></td>
</tr>
<tr>
<td>A. terreus (NIH 2624)</td>
<td>Broad Institute</td>
<td>Calhoun annotation system, including FGENESH; GENEID (Parra et al., 2000); and GENEWISE</td>
<td></td>
<td>GenBank: AAE000000000 <a href="http://www.broad.mit.edu/annotation/genome/aspergillus_group/">http://www.broad.mit.edu/annotation/genome/aspergillus_group/</a></td>
</tr>
<tr>
<td>A. flavus (NRRL 3357)</td>
<td>North Carolina State University/ Joint Craig Venter Institute (JCVI)</td>
<td>JCVI annotation pipeline including GlimmerHMM; Genezilla; SNAP; Genewise; Twinscan; PASA; and EvidenceModeler</td>
<td></td>
<td>GenBank: AAHF00000000 <a href="http://www.aspergillusflavus.org/genomics/">http://www.aspergillusflavus.org/genomics/</a></td>
</tr>
</tbody>
</table>
agricultural spoilage organism; *Aspergillus terreus*, sometimes an opportunistic pathogen, is also a source of lovastatin, one of the first of the hugely successful statins used therapeutically to inhibit cholesterol biosynthesis.

*A. nidulans* possesses a penicillin biosynthesis pathway similar to that found in the industrial producer *Penicillium chrysogenum*, and has most of the steps of the aflatoxin pathway found in *A. flavus* and *Aspergillus parasiticus*, so that it has become a key model system for studying these secondary metabolic pathways and their regulation (Bok et al., 2006b; Brakhage et al., 2004).

*A. nidulans* genetics has also contributed to eukaryotic cell biology beyond its close fungal relatives, examples being the discovery of γ-tubulin (Oakley and Oakley, 1989) and NudF, a homologue of the human lysencephaly protein (Xiang et al., 1995).

Since *A. nidulans* has played such an important role as a genetic model, Eurofungbase through the Eurofungbase Project decided to focus its community annotation efforts on *A. nidulans* in the first instance. We report the major findings of this exercise here, specific accounts on particular biological domains are presented in the following series of papers, and the newly annotated *A. nidulans* genome sequence is available in the CADRE database (Mabey et al., 2004) and will also be available in the Fungi section of ENSEMBL genomes at the EBI.

### 1.1.2. Annotation of aspergilli

The genome sequences of eight distinct *Aspergillus* species have been publicly released over the past 4 years. However, many of these genomes were annotated at different institutions using diverse methods over a relatively long time period, during which available tools and datasets have evolved rapidly. The inconsistency in annotation quality and completeness across these species hinders many avenues of comparative genomic research that depend on high-quality genome annotation, including evolutionary and functional studies (Wortman et al., 2006a).

The first three *Aspergillus* genome sequences, those for *A. fumigatus*, *A. oryzae* and *A. nidulans*, were published in 2005 and described in three companion papers (Galagan et al., 2005; Machida et al., 2005; Nierman et al., 2005). The *A. fumigatus* genome sequence (AF293) was generated through a collaboration between the Institute for Genomic Research (TIGR) and the Wellcome Trust Sanger Institute and deposited in GenBank; it has the accession AAH0000000000 (Nierman et al., 2005). The assembled genomic sequence was processed through the TIGR annotation pipeline, which subjected each sequence to a series of homology searches as well as algorithms for predicting genes (GlimmerM, Exonomy, Unveil, and GeneSplicer) (Majoros et al., 2003; Pertea et al., 2001). The gene prediction algorithms were trained with a limited dataset of *A. fumigatus* EST and cDNA sequences (Nierman et al., 2005) and the output of the pipeline was manually reviewed. An updated annotation for *A. fumigatus*, based on comparative genome data and involving targeted manual annotation, was released to GenBank in March, 2007 (Fedorova et al., 2008). The *A. oryzae* RIB40 genome, sequenced and analyzed by a consortium led by the National Institute of Advanced Industrial Science and Technology (AIST) in Japan, was released to DDBJ under the accession numbers AP001750 to AP007177. The AIST annotation pipeline incorporated EST and protein homology data via the gene predictors ALN (Gotoh, 2000), GeneDecoder (Asai et al., 1998), and GlimmerM. The pipeline was trained on a set of gene models which were constructed by alignment of known fungal proteins to the open reading-frames in the *A. oryzae* genome (Machida et al., 2005). *A. oryzae* annotation was last updated in DDBJ in December, 2005.

The *A. nidulans* genome (E. nidulans) genome, for strain FGSC A4, was sequenced and annotated by the Broad Institute of Harvard and MIT and submitted to GenBank and has the accession AACD00000000. The genome sequence was annotated using the Calhoun annotation system, which included protein homology searches and the gene prediction algorithms FGENESH (Salamov and Solovyev, 2000), FGENESH+, and GENEWISE (Birney et al., 2004). *A. nidulans* EST data was not incorporated into the gene predictions, but was used separately for validation (Galagan et al., 2005).

Comparison of these first aspergillus genome sequences revealed a surprising level of genetic variability. Proteome comparisons revealed an average amino-acid identity of less than 70% between each species pair, suggesting that they are as evolutionarily distant from each other as humans are from fish (Galagan et al., 2005b). Since these phylogenetic distances were so significant, it became clear that additional aspergilli would need to be sequenced to facilitate comparative analyses. More data would be needed to elucidate the specific gene differences and regulatory elements linked to the distinctive phenotypic and physiological properties important to the study of each organism. The proteome comparison also revealed the extent of gene model annotation differences between genomes, with the majority of identified orthologue groups (~80%) containing members differing in length and/or number of exons. Identified annotation problems included the merging of neighbouring loci, missed exon calls, and incorrect 5’ exons (Wortman et al., 2006a). A summary of the published genome sequences for *Aspergillus* species is given in Table 1, which provides information on the sequencing centres, software tools employed, and access to the sequences.

In addition to inconsistencies in gene model annotation caused by the lengthy timeframe over which the different genomes were sequenced and the diverse annotation processes employed at different sequencing centres, there are also significant differences in the functional annotation attached to gene products. While some groups will only attach a putative function based on a high stringency homology match to an experimentally characterized protein (Galagan et al., 2005), others have used more lenient criteria, employing resources such as InterPro (Mulder and Apweiler, 2007) and PFAM (Finn et al., 2008) profiles (Fedorova et al., 2008; Nierman et al., 2005) and the NCBI KOG (Tatusov et al., 2003) resource (Machida et al., 2005).

### 1.2. Eurofungbase

Eurofungbase is a coordination action programme funded by the European Commission under contract LSSG-CT-2005-018964. It comprises a community of 32 different partner laboratories in 11 European countries, supported by an Industry Platform of 13 companies. Eurofungbase aims to facilitate the construction of an integrated data warehouse to enable comparative and func-
tional genomic studies of filamentous fungi of scientific, medical, industrial, and agricultural importance. The Consortium has a number of different strategies for achieving its aims, but one of the most important approaches is to organise the manual annotation of the genomes of important model organisms by expert groups of researchers, and to consolidate and disseminate the results of such annotation exercises through journal publications and web facilities. It is within this context that Eurofungbase took on the task of re-annotating the A. nidulans genome with the help of colleagues from TIGR/J Craig Venter Institute and aspergillus research laboratories in the USA.

1.3. Evolution of A. nidulans genome data

The original public genome annotation of A. nidulans, described briefly above, consists of 9541 protein-coding gene predictions (Galagan et al., 2005). Each gene was assigned a unique locus identifier with the prefix AN followed by a four digit number between 0001 and 9541 and appended with the annotation version number 2 (e.g. AN0001.2). Version 1 was internal to the Broad Institute and not released widely. As of July, 2008, this is the version of the A. nidulans annotation that is still represented at GenBank, linked to accession AAC00006000, submitted in January, 2004. Of the genes overlapping EST alignments, approximately 70% were fully consistent with the EST data, while 30% showed some inconsistency (J. E. Galagan, personal communication). Comparative analysis with A. fumigatus and A. oryzae suggested that there were many neighboring loci inappropriately merged (Wortman et al., 2006a). Functional annotation was applied to gene products only when they exhibited high-identity matches to previously published, experimentally characterised, proteins within the fungal kingdom. This resulted in putative function assignments for approximately 3% of the predicted proteins.

In summer, 2005, NIAID requested that the annotation group at TIGR revisited the gene structure annotation of A. nidulans in advance of microarray design being planned by the NIAID-funded Pathogen Functional Genomics Resource Center (PFGRC). This re-annotation effort focused on the automated incorporation of EST data into the existing gene models, and the manual review and correction of merged loci. 32,931 EST and cDNA sequences compiled from GenBank and provided by C. d’Enfert and G. Goldman were aligned to the genome, and compared to the existing annotation using the PASA pipeline (Haas et al., 2003). These EST sequences collapsed to 8690 unique assemblies and were used to perform automated gene structure updates of 1146 genes. In addition, over 2000 genes that could not be computationally resolved with the current gene structure were manually reviewed and corrected on the basis of either protein homology or EST data. loci (494) were split into two or more distinct loci, and 214 new gene models were added. In addition, 426 gene models originally predicted by the Broad institute, but excluded from the earlier release because they did not meet minimum length criteria, were also incorporated. The final gene set consisted of 10,701 protein-coding gene predictions, with 4263 genes completely consistent with EST alignments. Locus identifiers were retained in all cases of one-to-one mapping (9447), whether the sequence changed or not, but the version number was incremented to 3 for all genes. Since the gene number was now over 10,000, there was a need to create new locus identifiers with five numeric digits after the AN prefix (e.g. AN10002.3). Functional annotation was supplied by the Broad Institute. As of July, 2008, this version 3 annotation data set is the current annotation reflected at the Broad Institute web site: http://www.broad.mit.edu/annotation/genome/aspergillus_group/

2. Eurofungbase community annotation

2.1. Main focus: gene function

The primary goal of the Eurofungbase annotation effort is to increase the numbers of A. nidulans proteins with informative functional assignments. Experts in various aspects of fungal and, particularly, A. nidulans biology were invited to participate in an ongoing annotation effort, which started with a jamboree in Autumn, 2006. Prior to this initial meeting, the version 3 protein sequences were subjected to a series of computational analyses intended to provide evidence for protein function. These included homology searches against the GenBank non-redundant database (nr), domain identification against the PFAM (Finne et al., 2008) and InterPro (Mulder and Apweiler, 2007) databanks, and the programs tmHMM (Sonnhhammer et al., 1998), which predicted transmembrane domains, and SignalP (Bendtsen et al., 2004), which predicts signal peptides.

To support the annotation effort, meeting participants were granted access to a compact summary of this computational evidence through the Manatee web-based interface (Wortman et al., 2006a). Manatee is a web-based manual annotation and analysis tool that acts as in interface between an underlying database and an annotator. Manatee’s interface allows the annotator to add, delete, and edit annotations attached to the protein, such as gene product names, gene symbols, EC, numbers and GO assignments. All such changes are referred back to the underlying database. Consortium members could continue to interact with the annotation database, through Manatee, on an ongoing basis for more than a year.

Over the course of the functional annotation effort, 2626 genes were reviewed and edited, with GO terms being added where possible and appropriate. Through this concerted effort, the percentage of A. nidulans gene products with an informative name has increased from approximately 3–19%. For the remaining un-reviewed genes, we have provided product names by transfer of information from A. fumigatus or A. niger orthologues, further increasing the proportion of gene products with informative names to 58%. As an indication of the changes arising from this project, we have also incremented the annotation version to 4 for all locus identifiers (e.g. AN****.4).

2.2. Assembly updates

2.2.1. Assembly of supercontigs and association with chromosomes

The initial Broad Institute assembly comprised 173 contigs, linked by end-sequenced BAC and fosmid bridges to form 16 supercontigs, corresponding to the 16 chromosome arms. A further 75 contigs were unplaced. Using BLAST (Altschul et al., 1990) and BL2seq (Tatusova and Madden, 1999) it was found that many of the Broad contigs overlapped so that 58 previously unassigned contigs could be incorporated into supercontigs, and the number of unsequenced gaps within supercontigs was reduced from 157 to 71 (see http://www.gla.ac.uk/ibls/molgen/aspergillus_contig-links.html and http://www.cadre-genomes.org.uk/Aspergillus_nidulans/Docs/revised_contig_overlaps.html). In five further cases, gaps are bridged by independently cloned sequences or retrotransposons with matching target-site duplications. The remaining gaps are assigned an arbitrary length of 1 kb in the revised assembly.

Supercontigs were related to chromosome arms using meiotically mapped and cloned linkage map markers (see http://www.gla.ac.uk/ibls/molgen/aspergillus/index.html). Over 185 such markers, identifiable with autocalled genes, are currently informative in this respect. With a few exceptions, most of which can be put down to reliance on inadequate linkage data, the correspondence between linkage and genome maps is excellent (Clutterbuck
and Farman, 2008). Four supercontigs have telomeric simple sequence (TTAGGG) repeats at one end, allowing the orientation of the arm to be validated independently of the genetic map, and using TERMINUS (Li et al., 2005, Clutterbuck and Farman (2008)) were able to identify telomeric simple sequences and subtelomeric contigs in the NCBI sequence trace archive and associate them with the remaining chromosome arms. While six different telomeres have identical telomeric contigs, the subtelomeric sequences are more varied, including a variety of simple sequences, and, in many cases, members of a specific class of telomere-linked helicases, all but one of which are truncated or degenerate (Clutterbuck and Farman, 2008). In most cases, the non-telomeric ends of supercontig arms were marked with clusters of A+T-rich, degraded transposable elements that are typical of centromeric DNA. In the revised assembly, centromeres, which have not been sequenced, are displayed as arbitrary gaps of 50 kb.

The one case in which there was a serious discrepancy between genomic and linkage maps concerned supercontig 6, associated with chromosome V. Here a telomeric link was found in the middle of the supercontig, requiring the splitting of this supercontig into 6a and 6b. New supercontig 6a was found to end in a short fragment of a ribosomal RNA repeat sequence and now spans a central region of chromosome arm V-L, between the nucleolus organizer and the centromere, the distal section of V-L being formed by supercontig 12. Supercontig 6b, starting with links to telomere T15, makes up chromosome arm V-R (see Fig 5.6 in Clutterbuck and Farman, 2008). The resulting rearranged supercontigs now correspond well with the linkage map of chromosome V, already known to include the nucleolus organizer (Brody et al., 1991), (approximate length 360 kb: Ganley and Kobayashi, 2007).

2.3. Assembly availability

In summary, the new assembly has 248 contigs, of which 231 contigs are assigned to 17 supercontigs (with 66 unsequenced gaps) that are mapped to eight linkages groups (approximately 30.5 Mb). This assembly is displayed within the Central Aspergillus Data Repository (CADRE) (Mabey et al., 2004). However, due to the recent improvements in assembly and annotation data, the current contigs no longer reflect the annotation within GenBank although the sequence remains the same. For this reason, contigs in CADRE are labeled with source identifiers from the Broad Institute (1.1–1.248), which coincide with the public sequence data within GenBank (AACD01000001 to AACD01000248). The supercontigs, annotated as scaffolds within GenBank (CH236920 to CH236935) have also changed. Therefore we have labeled the supercontigs in CADRE 1–16, with supercontig 6 replaced by 6a and 6b. These supercontigs are assigned to the linkages groups, which remain labeled as I–VIII.

2.4. Gene structure updates

The version 3 gene models, which were annotated on the 248 original contig sequences, were mapped to the revised chromosome sequences. The genome sequences were aligned using the NUCmer utility of the MUMmer package (Kurtz et al., 2004), and gene models were transferred on the basis of the resulting coordinate mapping. In this process, 82 gene predictions were unable to be transferred because of changes in the underlying sequence, and approximately 200 mapped gene pairs were found to overlap significantly. Gene (34) predictions were marked as putative pseudogenes, as the open reading-frames are disrupted in the reported genome sequence. The transferred gene models were the starting point for version 4 of the A. nidulans annotation.

During the course of functional annotation, consortium members with expertise in specific gene families were able to identify gene structures that were likely to be incorrect. They then submitted either virtual cDNA sequences representing the corrected gene structure, or protein sequences corresponding to corrected genes. PASA (Haas et al., 2003) was used to correct genes based on the submitted cDNA sequences, processing 74 gene structure updates, including 7 additional putative pseudogenes. Genewise (Binney et al., 2004) was used to instantiate gene structures based on the submitted protein sequences. These structures were reviewed manually and used to update 95 genes, including 3 additional putative pseudogenes, and to create 5 new gene models. The final gene count for the version 4 annotation is 10,605, which includes 63 putative pseudogenes. Two families of genes were significantly improved through these efforts. 39 of 119 cytochrome P450 genes were updated, as were 96 of the 342 predicted Zn(II)2Cys6 transcription factors.

This dataset reflects an iterative improvement in the A. nidulans predicted gene complement, but should not be considered a final product. The gene structure annotation of A. nidulans will continue to change over time as new experimental evidence and computational approaches arise. Of particular interest are recently published gene prediction programs that use machine-learning approaches to improve performance on eukaryotic genomes (Bernal et al., 2007; DeCaprio et al., 2007). The annotation process can also benefit from the incorporation of more diverse data types and predictions. One example is found in the companion article by Wang et al. FGB-08-96, which describes a new species-specific intron splice site predictor for aspergillus. In contrast to conventional gene finders producing entire gene structures, splice site predictors identify all candidate splice sites and assign probability values, information which can be used to correct gene structures or predict alternatively spliced genes.

3. New findings from the annotation

Subsequent articles in this issue do not aim to cover all the different families of genes, but rather reflect current research interests of the aspergillus community. Some of the genes included in the annotation (over 800) have been functionally characterised and/or genetically mapped prior to completion of the genome sequence, and these genes are mostly named according to the A. nidulans convention (Clutterbuck, 1973; Martinelli, 1994), in addition to the locus identifier in the AN*** format. Where genes have not been functionally characterised, the gene model name based on bioinformatics analysis is given only in the AN*** designation. For these uncharacterised genes, annotation is based on comparative genomics, and functions are suggested prior to functional analysis.

Since a genetic map of the eight linkage groups (chromosomes) of A. nidulans was produced prior to the genome sequencing project, it was instructive to compare the genetic and physical maps. In most cases, the correlation is good, and in the case of chromosome V, the linkage map assisted in the correct assembly of the genome sequence.

3.1. Transposable elements

Transposable elements (TEs) were catalogued mainly by V.V.Kapitonov & J.Jurka, Genetic Information Research Institute, and can be found in Repbase (http://girinst.org/repbase/). 1275 insertions of whole or fragmented elements were identified, representing 14 retroposons and 19 DNA transposon families. Their sequences and distribution are described elsewhere (Clutterbuck et al., 2008). A total of 306 autochorted genes are involved with TEs identified by Kapitonov & Jurka. 151 are found wholly within TEs and many of these have recognizable transposition-related fea-
tatures. 103 genes overlap TEs or TE fragments, and 52 others have small TEs or fragments within them, only 14 of these being wholly within introns. In many cases the TEs are affected by RIP, and their associated autocalleed genes have atypical (probably unrealistic) structures, often including multiple long introns, e.g. the mean length of 67 introns in 24 genes overlapping Mariner-6 elements was 152 bp: three times the modal intron length.

In only two cases have the affected genes been previously identified: the 3’ end of AN7818.3 (stcf) overlaps a degraded Mariner-4 AN copy by just three nucleotides. Secondly, (Cultrone et al., 2007) have reported a case where a non-autonomous Helitron-N1_AN element has been created as a result of a 3’ deletion of an autonomous Helitron1-AN element and readthrough into the xanA gene (AN10081.3). The new element has transposed once to give a second copy of the xanA promoter and 5’ region, named pxA (AN11581.3). This pseudogene is transcribed, but the resulting mRNA is rapidly destroyed by nonsense-mediated decay.

3.2. Transcription factors

Eukaryotic transcription factors can be classified on the basis of their characteristic DNA-binding domains. The A. nidulans genome was analyzed for the presence of twelve different classes of DNA-binding transcription factors (Table 2). DNA-binding proteins without transcription factor activity were not included in the analysis. To identify putative transcription factors in the A. nidulans genome, BlastP searches were performed using functional characterized transcription factors and/or predicted transcription factors from Saccharomyces cerevisiae and Neurospora crassa as queries. Predicted proteins that showed significant similarity to the above-mentioned transcription factors, but lacked the typical DNA-binding domain, were subsequently checked manually for alternative gene models that would include a DNA-binding domain. An overview of the number of putative transcription factors in the A. nidulans genome is given in Table 2 and further details are given as supplementary data (Suppl. Tables 1–13).

As shown in Table 2, the family of Zn(II)2Cys6 transcription factors (MacPherson et al., 2006) is the largest, consisting of 330 proteins. The number of annotated proteins belonging to this class has been increased significantly from the initial annotation (Galagan et al., 2005) by improving the gene models manually. By addition of 5’exons, and/or changing intron/exon boundaries, in many cases a complete conserved six cysteine motif (CX2CX2CX2CX3-CX3-CX6-C) that was absent in a large number of the original gene models could be included in the protein sequence. For only a small number of putative Zn(II)2Cys6 TF proteins were we unable to identify a specific DNA-binding domain even by manual annotation (Suppl. Table 13).

Members of an interesting subclass of Zn(II)2Cys6 transcription factors contain a C2H2 DNA-binding domain in addition to the Cys6 motif. A total of 12 such transcription factors have been identified (Supp. Table 2). The function of those genes has not been studied in A. nidulans. In Colletotrichum lagenarium (Cmr1) and Magnaporthe grisea (Pig1) the double motif transcription factors are involved in the regulation of mycelial melanin biosynthesis (Tsuji et al., 2000).

It is a major challenge for future work to perform functional analysis of the members of the different families of transcription factors. High-throughput deletion strategies or overexpression approaches are useful methods for studying their function, as has proved to be the case for N. crassa (Colot et al., 2006) For A. nidulans, only 21 Zn(II)2Cys6 transcription factors have been functionally analyzed in more detail (Suppl. Table 1).

Previous work has indicated that, in various Aspergillus species, transcription factors are often clustered in the genome with their target genes (Cazelle et al., 1998; Felenbok et al., 2001; Gomi et al., 2000; Hull et al., 1989; Lamb et al., 1990; Unkles et al., 1992; Woloshuk et al., 1994; Yuan et al., 2008a). This applies, for example, to genes associated with primary and secondary metabolism and carbohydrate metabolism. Based on this finding, analysis of genes adjacent to the Zn(II)2Cys6 transcription factors might give information about their possible target genes and, hence, function (Flipphi et al., 2006; Ram and Punt, unpublished results).

In addition, phylogenetic analyses of the protein sequences of Zn(II)2Cys6 transcription factors may be useful for the identification of sub-clusters with related functions. In a simple phylogenetic analysis of the 330 A. nidulans Zn(II)2Cys6 transcription factors, approximately 56 sub-clusters (defined as proteins with a reciprocal BlastP e-value of <−20 towards each other) could be identified. One of these clusters contained both the AmyR (AN2016.4) and InuR (AN3835.4) transcription factors involved in the regulation of enzymes degrading starch and inulin, respectively (Yuan et al., 2008a). The possible roles in polysaccharide catabolism of other transcription factors in this specific subcluster (AN7343.4, AN7346.4, AN8596.4, and AN10423.4) await further analysis.

3.3. Polysaccharide degrading enzymes

A. nidulans is found naturally in the rhizosphere, where it is able to utilise decaying plant material by means of secreted protein-and polysaccharide-degrading enzymes. Despite significant differences in the set of putative plant polysaccharide-degrading enzymes predicted from the genome sequence of A. niger, A. nidulans, and A. oryzae, no large differences were found in their growth on cellulose, xylan, pectin or galactomannan (Coutinho et al. FGB-08-108). The apparent differences in the repertoire of en-

---

**Table 2**

Overview of the twelve annotated classes of transcription factors (TF) in the A. nidulans genome.

<table>
<thead>
<tr>
<th>TF class</th>
<th>Table in this manuscript</th>
<th>DNA binding-domain</th>
<th>pfam Domain</th>
<th>number of genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Supplementary Table 1</td>
<td>Zn2Cys6</td>
<td>00172</td>
<td>330</td>
</tr>
<tr>
<td>2</td>
<td>Supplementary Table 2</td>
<td>C2H2+Zn2Cys6</td>
<td>00096 &amp; 00172</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>Supplementary Table 3</td>
<td>C2H2</td>
<td>00096</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>Supplementary Table 4</td>
<td>GATA</td>
<td>00320</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>Supplementary Table 5</td>
<td>b-ZIP</td>
<td>07716</td>
<td>22</td>
</tr>
<tr>
<td>6</td>
<td>Supplementary Table 6</td>
<td>b-HLH</td>
<td>00010</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>Supplementary Table 7</td>
<td>CBF/CCAAT</td>
<td>00800</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>Supplementary Table 8</td>
<td>APSES</td>
<td>02292</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>Supplementary Table 9</td>
<td>Myb-like</td>
<td>00249</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>Supplementary Table 10</td>
<td>Fork head domain</td>
<td>pfam00250 and 00498</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>Supplementary Table 11</td>
<td>Homeodomain-box</td>
<td>00046</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>Supplementary Table 12</td>
<td>MADS-box</td>
<td>00319</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>Supplementary Table 13</td>
<td>Miscellaneous Zn2Cys6</td>
<td></td>
<td>15</td>
</tr>
</tbody>
</table>
zymes may, therefore, reflect specific adaptations to the natural biotope of the individual species. The identification of transcription factor (TF) target sites within the promoter regions of genes encoding polysaccharide-degrading enzymes is rarely helpful in predicting function. Thus while there is a correlation between the presence of XlnR sites and xylose induction for A. niger genes, this is not strictly true for the other two species; it is likely that TF binding sites in the promoters vary between Aspergillus species.

3.4. Primary metabolism

Primary metabolism and its regulation has been a major pre-genomic research area, and the post-genomic annotation has now identified genes for most of the central pathways. Several of these pathways, including glycolysis, the pentose phosphate pathway, TCA cycle, and the \( \alpha \)-arabinose/\( \alpha \)-xylose oxidoreductive pathway exhibit multiple genes for some individual steps (Flippghi et al. FGB-08-100). Phylogenetic analysis shows that these gene duplications were early events in the evolution of the aspergilli and it can be assumed that they have aided their proliferation in specific habitats. In addition to the pyruvate dehydrogenase and 2-oxo-glutarate dehydrogenase complexes, a third 2-oxo-dehydrogenase complex was discovered, and this is absent from yeast. The loci involved in this third complex are AN1726, AN3639, and AN8559. This enzyme is likely to be involved in the degradation of branched-chain aliphatic amino acids valine and isoleucine (like the orthologous complex in humans). A gene encoding glucose oxidase was found, suggesting an oxidative catabolic pathway for glucose. However, A. nidulans appears to lack the hydrolysing enzyme to linearise glucono-lactone. Moreover, since mutual disruption of the function of the genes for hexokinase (\( \text{hxkA} \)) and glucokinase (\( \text{glkA} \)) effectively abolishes growth on glucose (Flippghi et al., 2003), it is unlikely that glucose oxidation could lead to growth. Interestingly, A. nidulans appears distinct from all other aspergilli in lacking inositol-phosphate phosphatase – catalyzing the last step of inositol biosynthesis – which is non-essential in S. cerevisiae (Lopez et al., 1999). The encoding gene is found in synteny with its step of inositol biosynthesis – which is non-essential in lacking inositol-phosphate phosphatase – catalyzing the last genes, AN9409 and AN9410, while in the other seven public Aspergillus genomes, the structurally well-conserved inositol biosynthesis have been the subject of substantial previous study, in functional detail. Similarly, the pathways of fungal aminoacid biosynthesis have been the subject of substantial previous study, including use as nutritional markers, and many could also be annotated.

Blast analysis identified signatures in several other genes which suggested putative roles in nitrogen metabolite transport, assimilation or aminoacid biosynthesis. These included a putative amino-acid permease (AN7392), purine transporter (AN7955), an additional acetamidase (AN9138) and four possible additional members of the arginase family (AN7488, AN6869, AN7669, and AN3965). Bioinformatic analysis predicted functions for AN7488 and AN6869 as agmatinases and thus roles in polynucleoside biosynthesis. During searches for genes with amino-acid biosynthetic function, several aromatic aminotransferases were identified (AN6338, AN5041, AN8172, and AN4156), at least one of which is a candidate for the final step in tyrosine biosynthesis. In addition, homology to the histidinol dehydrogenase region of the tri-functional hisDEI (AN0797) was identified in AN2723, but the other domains of this protein were absent. Functional analysis will be required to determine whether these genes indeed have roles in nitrogen assimilation, amino-acid metabolism or perhaps in secondary metabolism.

3.6. Sulphur metabolism

Almost all known genes related to eukaryotic sulphur metabolism have been annotated, and assigned an appropriate function. Within the last 2 years, the \( \text{astA} \) gene encoding an alternative sulphur transporter has been identified and characterized. The gene's expression is strongly regulated by sulphur metabolite repression, which suggests that \( \text{astA} \) protein is a specific sulphate transporter. The protein is distinct from known sulphate permeases and belongs to a poorly characterized large Dal5 family of allantoate transporters. Interestingly, \( \text{astA} \) appears to be a nonfunctional gene in some strains of A. nidulans. Its orthologues are present in only a few fungal species, in particular in plant pathogenic/saprophytic fungi (Piłsyk et al., 2007).

A group of genes encoding enzymes directly or indirectly implicated in homocysteine metabolism are induced by this amino acid. Coordinated regulation of this “homocysteine regulon” (Sienko et al., 2007) may prevent homocysteine reaching toxic concentrations within the cell. Recent evidence suggests that this response to homocysteine may be part of a general reaction to stress that also involves the unfolded protein response. While the molecular mechanisms involved in regulation of the “homocysteine regulon” remain to be elucidated, it is clear that the response is not part of the sulphur metabolite repression system (SMR), which controls a number of sulphur-related genes, particularly those implicated in sulphate assimilation.

3.7. Secondary metabolism

Secondary metabolism is the hallmark of most filamentous ascomycetes, and one of the intriguing findings arising from the genome sequencing projects is how few orthologous genes and gene clusters are shared between the different Aspergillus species (Nierman et al., 2005). The functions of most of the new secondary metabolic clusters revealed by the genome project remain unknown, and their biosynthetic products cannot be predicted from sequence alone. Nevertheless, the availability of the genome sequence has stimulated functional analysis studies. This has already led to the discovery of a number of metabolic products not previously reported for A. nidulans, namely asparaginones A and B (Bergmann et al., 2007), terrerquinone A (Rok et al., 2006a) and the emericellamides (Chiang et al., 2008). An analysis of predicted
non-ribosomal peptide synthetase genes has been carried out (von Döhren FGB-08-122).

3.8. Cytochrome P450

The extensive metabolic capacity of A. nidulans is also reflected in the 119 cytochrome P450 (CYP) genes (including 8 pseudo-genomes) that have now been annotated (Kelly et al. FGB-08-98), both manually and by exploiting the query functions of the e-Fungi data warehouse (Cornell et al., 2007). The functions of 13 of these have been determined to date, including genes for ergosterol and dityrosine biosynthesis. Genes (32) were located close to known secondary metabolic genes such as those encoding nonribosomal peptide synthetases and polyketide synthases, probably reflecting their role in decorating secondary metabolites. Also identified were 8 putative NADPH cytochrome P450 reductase sequences. The protein encoded by AN0595 on chromosome VIII has 91% amino acid sequence identity to the CprA protein from A. niger, 88% identity to the CprA from A. fumigatus, and is the most probable candidate reductase for the majority of the CYPs.

3.9. Peroxisomes

Fungal microbodies (peroxisomes, Woronin bodies) are inducible organelles that proliferate or are degraded (via an autophagy-related process termed pexophagy) in response to nutritional cues. Proteins involved in microbody biogenesis/proliferation are designated peroxins and are encoded by PEX genes, while genes involved in autophagy (including pexophagy) are known as ATG genes. A reappraisal of the A. nidulans genome for the presence of PEX and ATG genes has identified a number of previously missed genes. This analysis has led to the conclusion that the basic set of genes involved in microbody biogenesis, proliferation, and degradation are conserved in A. nidulans (Kiel and van der Klei FGB-08-78). The major differences between filamentous ascomycetes like A. nidulans and other organisms appears to be an aberrant RING structure on the peroxin Pex2p (part of the so-called “importomer”), the presence of a Pex16p orthologue (previously identified mainly in higher eukaryotes) and the novel protein Pex14/17p (related to the yeast-specific peroxin Pex17p), as well as multiple Pex11p paralogues (a feature seen mainly in higher eukaryotes). With respect to pexophagy, a receptor protein required to link microbodies destined for degradation to the autophagy machinery could not be identified in A. nidulans. Nevertheless, A. nidulans contains an Atg11p orthologue, a protein associated with selective autophagy, implying that selective microbody degradation occurs in this filamentous fungus. Because both microbody biogenesis/proliferation and autophagy/pexophagy have features that more closely resemble organelle formation/degradation in mammals rather than yeast, filamentous fungi like A. nidulans are ideal model systems of peroxisome homeostasis in man.

3.10. Cell wall genes

The hyphal walls of A. nidulans consist of an internal electron-transparent layer surrounded by an electron-dense outer layer (Jeong et al., 2004). This is consistent with the notion that the internal layer mainly contains stress-bearing polysaccharides, whereas the outer layer is enriched with glycoproteins (De Groot et al., 2005). The inventory of cell wall genes reveals the presence of genes encoding polysaccharide synthases, including genes responsible for the production of the stress-bearing polysaccharides 1,3-α-glucan, 1,3-α-glucan, 1,3-β-glucan, and chitin (De Groot et al. FGB-08-116). In addition, A. nidulans contains a gene (AN8444, designated as celA) of the product of which shows similarity to bacterial plant and cellulose synthases and is predicted to be responsible for the production of 1,3-β-/1,4-β-glucan. This agrees with the presence of a linear 1,3-β-/1,4-β-glucan in the alkali-insoluble fraction of the hyphal wall of A. fumigatus (Fontaine et al., 2000). Importantly, four predicted GPI-protein-encoding genes were identified that are believed to have transglucosidase activity, and to act on 1,3-α-glucan (AN3308, AN4507, and AN6324) or on 1,4-β-glucan (AN2385) (van der Kaaij et al., 2007; Yuan et al., 2008b). Whereas many GPI-proteins are predominantly located in the plasma membrane, others are incorporated in the fungal cell wall (reviewed in (De Groot et al., 2005)). Mass spectrometric analysis of a tryptic digest obtained by ‘cell wall shaving’ (Yin et al., 2008) showed the presence of twelve proteins, including ten predicted GPI-proteins. Hydrophobins are small amphipathic proteins that are only found in mycelial species and are deposited on the cell surface of conidia/ascospores, basidioecars and aerial hyphae to render them hydrophobic (Wösthen, 2001). Six putative hydrophobin-encoding genes were identified in the A. nidulans genome, including the already known rodA and dewA. Finally, some genes involved in spore wall maturation were identified, including some encoding chitin deacetylases, which are predicted to convert chitin into chitosan. In addition, two predicted cytosolic proteins were found that show similarity to the ScDit proteins. These synthesize precursors of the dityrosine-containing macromolecule found in the outer wall layer of the ascospores of S. cerevisiae (Briza et al., 1994).

3.11. Secretion

The secretion of proteins is important to the natural lifestyles of the aspergilli and becomes critical when those species are exploited as hosts for the commercial secretion of heterologous proteins. Their natural lifestyles make the aspergilli more effective secretors than the model fungus S. cerevisiae, suggesting that the aspergilli might have a better capacity for secretion of proteins in general. That is confirmed by the experience of the biotechnological industry, although S. cerevisiae can be improved to secrete some proteins to commercially acceptable yields. Therefore, the main question facing the annotation team was: does the genome annotation provide any clues as to why the aspergilli might be better protein secretors than the model fungus S. cerevisiae and might they be better suited to express proteins with human-like glycosylation? The annotation alone does not provide simple answers to these questions (Geyens et al. FGB-08-112; Tsang et al. FGB-08-121). The core facets of the chaperone cycle, formation of disulphide bonds and the unfolded protein response (UPR) show a high degree of similarity between species. Even so, differences between the aspergilli and the yeasts can be seen in the nucleotide exchange reaction within the chaperone cycle, the protein disulphide isomerase protein family and the activation mechanisms of the UPR mediated by the Hac bZip transcription factor. The functional consequences of these differences remain to be assessed.

In eukaryotes, N-glycosylation is performed via the transfer of a lipid-linked precursor structure (GlClManbGlCNAC2) onto a nascent polypeptide chain within the lumen of the ER. In the aspergilli, orthologues were identified for almost all genes involved in the generation of activated sugar donors as well as the synthesis and transfer of the precursor structure to a protein backbone. Also, the genes involved in the ER-processing of protein-linked N-glycans seem to be present. However, differences were observed with S. cerevisiae and in some cases this makes the aspergillus system more like that of mammals, e.g. the presence of a reglucosylation enzyme (UGGT), involved in the calnexin-based glycoprotein quality control. In general, the N-glycans on proteins secreted by the aspergilli are of the high-mannose type, sometimes decorated with monosaccharides such as GalF. In contrast to S. cerevisiae, the aspergilli barely synthesize hyperglycosyl structures. Nevertheless, they
possess several orthologues of yeast mannosyltransferases that are known to be involved in the process of hypermannosylation. Several orthologues for mannosidases of family 47 and 92 were identified but it remains to be evaluated whether they act upon N-glycans inside (as in the mammalian Golgi apparatus) or outside the cell. Golgi-based demannosylation would again provide the aspergillus glycosylation pathway with mammalian similarity.

3.12. Polar growth

The signaling pathways involved in morphogenesis and calcium responses are reasonably well-conserved when compared to those of the well-studied yeast models (Harris et al. FGB-08-135). However, A. nidulans possesses many additional genes implicated in morphogenesis, development, and calcium signaling that are not found in yeast. The function of most of these genes remains a mystery, though it seems likely that they will contribute to unique aspects of hyphal growth (i.e., highly polarized growth) and development (conidiation, cleistothecium formation). Notably, A. nidulans shares with migratory animal cells and neurons several genes involved in actin organization that are otherwise absent from yeast. The number of genes potentially implicated in hyphal morphogenesis is therefore rather large and may exceed initial expectations.

4. Data availability

All data generated from this project – refined assembly, gene structures and annotations - have been deposited within CADRE (http://www.cadre-genomes.org.uk). CADRE (Central Aspergillus Data Repository) is a public resource that provides web-based tools for visualising and analysing genomic features identified within aspergillli. These tools offer simple displays for viewing annotation assigned to predicted genes (e.g. gene symbol, public loci and GO terms) and to their protein products (e.g. family and domain similarity matches), as well as complex displays for viewing gene and other features (e.g. repeated sequences) in the context of an assembly. Using the customised search facilities provided by CADRE, all A. nidulans genes can be easily sought and identified by their unique public locus identifier (AN****.4).

Acknowledgements

We acknowledge financial support by the European Commission under contract LSGC-CT-2005-018964. MC, and the use of the e-Fungi data warehouse, was supported by a grant to SGO and others as part of the BBSC’s Bioinformatics and e-Science programme II. We wish to thank Dr Michael Anderson for his input on the assembly, whilst at The University of Manchester. We also acknowledge Todd Creasy for Manatee set-up and support, currently at IGS; Brian Haas for data management and computational support (annotation), currently at the Broad; Joshua Orvis for data management and computational support (annotation), currently at IGS; Jonathan Crabtree for data management and computational support (orthologues), currently at IGS. Sandra MJ. Langeveld is acknowledged for her assistance in preparing part of the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.fgb.2008.12.003.

References


