



Universiteit
Leiden
The Netherlands

Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88

Pel, H.J.; Winde, J.H. de; Ram, A.F.J.; Hondel, C.A.M.J.J. van den; et al

Citation

Pel, H. J., Winde, J. H. de, Ram, A. F. J., & Hondel, C. A. M. J. J. van den. (2007). Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nature Biotechnology*, 25(2), 221-231. doi:10.1038/nbt1282

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/67447>

Note: To cite this publication please use the final published version (if applicable).

Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88

Herman J Pel¹, Johannes H de Winde^{1,2}, David B Archer³, Paul S Dyer³, Gerald Hofmann⁴, Peter J Schaap⁵, Geoffrey Turner⁶, Ronald P de Vries⁷, Richard Albang⁸, Kaj Albermann⁸, Mikael R Andersen⁴, Jannick D Bendtsen⁹, Jacques A E Benen⁵, Marco van den Berg¹⁰, Stefaan Breestraat¹, Mark X Caddick¹¹, Roland Contreras¹², Michael Cornell¹³, Pedro M Coutinho¹⁴, Etienne G J Danchin¹⁴, Alfons J M Debets¹⁵, Peter Dekker¹, Piet W M van Dijk¹, Alard van Dijk¹, Lubbert Dijkhuizen^{16,17}, Arnold J M Driessen¹⁷, Christophe d'Enfert¹⁸, Steven Geysens¹², Coenie Goosen^{16,17}, Gert S P Groot¹, Piet W J de Groot¹⁹, Thomas Guillemette²⁰, Bernard Henrissat¹⁴, Marga Herweijer¹, Johannes P T W van den Hombergh¹, Cees A M J J van den Hondel²¹, Rene T J M van der Heijden²², Rachel M van der Kaaij^{16,17}, Frans M Klis¹⁹, Harrie J Kools⁵, Christian P Kubicek²³, Patricia A van Kuyk²¹, Jürgen Lauber²⁴, Xin Lu²⁵, Marc J E C van der Maarel¹⁶, Rogier Meulenberg¹, Hildegard Menke¹, Martin A Mortimer¹¹, Jens Nielsen⁴, Stephen G Oliver¹³, Maurien Olsthoorn¹, Karoly Pal^{15,26}, Noël N M E van Peijl¹, Arthur F J Ram²¹, Ursula Rinas²⁵, Johannes A Roubos¹, Cees M J Sagt¹, Monika Schmoll²³, Jibin Sun²⁵, David Ussery²⁷, Janos Varga^{26,28}, Wouter Vervecken¹², Peter J J van de Vondervoort²¹, Holger Wedler²⁴, Han A B Wösten⁷, An-Ping Zeng²⁵, Albert J J van Ooyen¹, Jaap Visser²⁹ & Hein Stam¹

The filamentous fungus *Aspergillus niger* is widely exploited by the fermentation industry for the production of enzymes and organic acids, particularly citric acid. We sequenced the 33.9-megabase genome of *A. niger* CBS 513.88, the ancestor of currently used enzyme production strains. A high level of synteny was observed with other aspergilli sequenced. Strong function predictions were made for 6,506 of the 14,165 open reading frames identified. A detailed description of the components of the protein secretion pathway was made and striking differences in the hydrolytic enzyme spectra of aspergilli were observed. A reconstructed metabolic network comprising 1,069 unique reactions illustrates the versatile metabolism of *A. niger*. Noteworthy is the large number of major facilitator superfamily transporters and fungal zinc binuclear cluster transcription factors, and the presence of putative gene clusters for fumonisins and ochratoxin A synthesis.

¹DSM Food Specialties, PO Box 1, 2600 MA Delft, The Netherlands. ²Kluyver Centre for Genomics of Industrial Fermentation, Department for Biotechnology, Delft University of Technology, Julianalaan 67, 2628 BC Delft, The Netherlands. ³School of Biology, University of Nottingham, University Park, Nottingham, NG7 2RD, UK. ⁴Center for Microbial Biotechnology, Technical University of Denmark, Building 223, Søtofts Plads, DK-2800 Kgs. Lyngby, Denmark. ⁵Section Fungal Genomics, Laboratory of Microbiology, Wageningen University, Dreijenlaan 2, 6703 HA Wageningen, The Netherlands. ⁶Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, UK. ⁷Microbiology, Science Faculty, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. ⁸Biomax Informatics AG, Lochhamer Str. 9, 82152 Martinsried, Germany. ⁹CLC bio, Gustav Wieds Vej 10, 8000 Aarhus C, Denmark. ¹⁰DSM Anti Infection, PO Box 425, 2600 KA Delft, The Netherlands. ¹¹The University of Liverpool, School of Biological Sciences, Biosciences Bld., Crownstreet, Liverpool L69 7ZB, UK. ¹²Ghent University and VIB, Dept. Molecular Biomedical Research, Unit of Fundamental and Applied Molecular Biology, Technologiepark 927, 9052 Gent, Belgium. ¹³Centre for the Analysis of Biological Complexity, Faculty of Life Sciences, The University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK. ¹⁴Architecture et Fonction des Macromolécules Biologiques, UMR6098, CNRS. Universités Aix-Marseille I & II, Case 932, 163 Avenue de Luminy, 13288 Marseille, France. ¹⁵Laboratory of Genetics, Wageningen University, Arboretumlaan 4, 6703 BD Wageningen, The Netherlands. ¹⁶Centre for Carbohydrate Bioprocessing (CCB) TNO-RuG, PO Box 14, 9750 AA Haren, The Netherlands. ¹⁷Department of Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute and Materials Science Center Plus, University of Groningen, PO Box 14, 9750 AA Haren, The Netherlands. ¹⁸Unité Postulante Biologie et Pathogénicité Fongiques, INRA USC 2019, Institut Pasteur, 75724 Paris Cedex 15, France. ¹⁹Swammerdam Institute for Life Sciences, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands. ²⁰Laboratoire de Microbiologie, UMR 77, Pathologie Végétale, Université d'Angers, 49045 Angers Cedex, France. ²¹Institute of Biology Leiden, Leiden University, Molecular Microbiology, Wassenaarseweg 64 2333 AL Leiden, The Netherlands. ²²CMBI, Radboud University Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands. ²³Institute of Chemical Engineering, Research Area Gene Technology and Applied Biochemistry, Technical University Vienna, 1060 Vienna, Austria. ²⁴Qiagen Genomics Services, Qiagen GmbH., 40724 Hilden, Germany. ²⁵Helmholtz Center for Infection Research (former GBF-German Research Centre for Biotechnology), Inhoffenstrasse 7, D-38124 Braunschweig, Germany. ²⁶Department of Microbiology, Faculty of Sciences, University of Szeged, PO Box 533, H-6701 Szeged, Hungary. ²⁷Center for Biological Sequence Analysis, BioCentrum-DTU, Building 301, The Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark. ²⁸CBS Fungal Biodiversity Centre, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. ²⁹Fungal Genetics and Technology Consultancy, PO Box 396, 6700 AJ Wageningen, The Netherlands. Correspondence should be addressed to H.S. (hein.stam@dsm.com).

Received 17 July 2006; accepted 27 November 2006; published online 28 January 2007; doi:10.1038/nbt1282

Aspergillus niger, a member of the black aspergilli, is widely used in biotechnology for the production of food ingredients, pharmaceuticals and industrial enzymes. In their natural habitat *A. niger* strains secrete large amounts of a wide variety of enzymes needed to release nutrients from biopolymers. This high secretory capacity is exploited by industry in both solid state and submerged fermentations^{1,2}. *A. niger* has a long tradition of safe use in the production of enzymes and organic acids. Many of these products have obtained GRAS (generally regarded as safe) status³. *Aspergillus* enzymes are used in starch processing, baking, brewing and beverage industries, in animal feed and in the paper and pulping industry. Furthermore, *A. niger* is used as host for the production of heterologous proteins^{4,5} and as cell factory for the production of citric acid and gluconic acid⁶. *A. niger* exhibits a remarkably versatile metabolism, enabling growth on a wide range of substrates under various environmental conditions. Its ability to degrade a range of xenobiotics through various oxidative, hydroxylation and demethylation reactions provides potential for use in bioremediation⁷.

In this paper we describe the genomic DNA sequence of *A. niger* strain CBS 513.88, its annotation and an initial gene expression study using Affymetrix DNA microarrays. CBS 513.88 is an early ancestor of currently used enzyme production strains. Several features are compared with the recently published genomes of the aspergilli *A. oryzae*⁸, *A. nidulans*⁹ and *A. fumigatus*¹⁰. The availability of the genome will facilitate the development of new products, improved strains and more efficient processes.

RESULTS

Genome sequence and analysis

The genome of *A. niger* CBS 513.88 was sequenced using an ordered set of large insert *Escherichia coli* bacterial artificial chromosomes (BACs) in a process called BAC walking. First, a large insert (up to 150 kb) BAC library covering the genome more than 20 times was constructed and used to generate over 18,000 BAC end sequences with an average length of 500 base pairs (bp). In parallel, several of the larger BACs were sequenced using a 7.5× coverage shotgun approach. Subsequently, sequences of these fully sequenced 'seed' BACs were compared to the BAC ends, allowing the selection of a new set of BACs showing overlap with one of the sequenced BACs. The newly selected BACs were used for shotgun sequencing. A total of 505 BACs, representing a minimal tiling set covering the genome, were selected and sequenced using the BAC walking approach. The assembled genome sequence consists of 468 DNA contigs spanning a total of 33.9 million unique bp (Mbp) arranged in 19 supercontigs. Recently an 8.9× shotgun coverage draft sequence of the citric acid-producing *A. niger* ATCC 1015 was made available by the DOE Joint Genome Institute (<http://www.jgi.doe.gov/aspergillus>). Based on Blastn analysis the ~37.2-Mbp genome sequence is 97% identical to and largely colinear with the genome sequence published here.

Gene identification and annotation

Algorithms specifically trained on the basis of genes known in *A. niger* and related filamentous fungi were used to predict 14,165 protein coding genes (Table 1). Because automated gene-model predictions frequently fail to correctly predict all intron and exon boundaries, extensive manual curation was used to verify the predicted gene models and to assign gene functions on the basis of similarity to known proteins with an established function. In this way we improved mainly the models of genes encoding proteins with similarity to known proteins.

Table 1 Genome characteristics of *A. niger* CBS 513.88

Size assembled genome (Mb)	33.9
GC content assembled genome (%)	50.4
Protein-coding genes (all)	14,165
Protein-coding genes (<100 aa)	927
GC content protein-coding genes (%)	53.4
Average gene length (bp)	1,572
Gene density (genes/kb)	0.42
Average number of introns per gene	2.57
Average intron size (bp)	97
Average exon size (bp)	370
Number of tRNA genes	269

The specificity of the 269 tRNA genes and the codon usage are presented in **Supplementary Tables 1 and 2** online and confirm the published usage¹¹. *A. oryzae* has a comparable set of tRNAs (245), whereas *A. nidulans* and *A. fumigatus* contain 188 and 179 tRNAs, respectively. Whereas the observed tRNA density is higher for *A. niger* and *A. oryzae*, the codon distribution is similar for all four species. The functional catalog (FunCat) classification system¹² was applied to functionally describe the *A. niger* predicted proteome. Of the 14,165 predicted proteins, approximately half (6,505) could be assigned to functional protein classes relating to metabolism, cellular transport and protein fate (Fig. 1). A detailed comparison was made between *A. niger* and eight other filamentous fungi (Fig. 1): 9,253 *A. niger* proteins have an ortholog in at least one of the nine queried species, 1,992 proteins have orthologs in all species, whereas 3,373 proteins are shared between the filamentous fungi. The distribution of orthologous proteins over the various FunCat classes is shown (Fig. 1). Compared to the other filamentous fungi *A. niger* contains a remarkably large number of unique proteins involved in C-compound, carbohydrate, lipid, fatty-acid and isoprenoid metabolism and secondary metabolism, reflecting the versatility of *A. niger* as a cell factory. In contrast, the number of unique proteins involved in cellular transport, protein secretion and fate are rather constant between the various filamentous fungi. Apparently *A. niger* is able to use its secretion machinery in a very efficient way.

Phylogenetic relationship and synteny between aspergilli

Phylogenetic relationships between filamentous fungi have often been based on ribosomal DNA sequences or single-gene families¹³. We selected twenty strictly orthologous sequences from *A. niger* and eight filamentous fungi. For *A. terreus* and *A. clavatus* no annotated proteome is available, hence translations of orthologous gene fragments were used instead. The multiple sequence alignment was manually curated and used to build a maximum likelihood tree (Supplementary Table 3 and Supplementary Fig. 1 online). In this tree *A. niger* is closely related to both *A. terreus* and *A. oryzae* and separate from *A. fumigatus* and *A. clavatus*, which have smaller genomes (29–30 Mb), and *A. nidulans*, which branches earlier.

Of the 8,695 *A. niger* genes that have an ortholog in *A. nidulans*, *A. oryzae* and *A. fumigatus*, 6,755 (78%) show conservation of neighboring orthologs (synteny) in at least one of the three other species; 4,189 genes (48%) are syntenic in all four *Aspergillus* species and were plotted on the proposed physical map of *A. niger*. Large parts of the *A. niger* genome show a high conservation of gene order with one of the other *Aspergillus* species, in line with previous observations⁹. However, within those regions numerous intrachromosomal rearrangements and microinversions have occurred (Fig. 2).

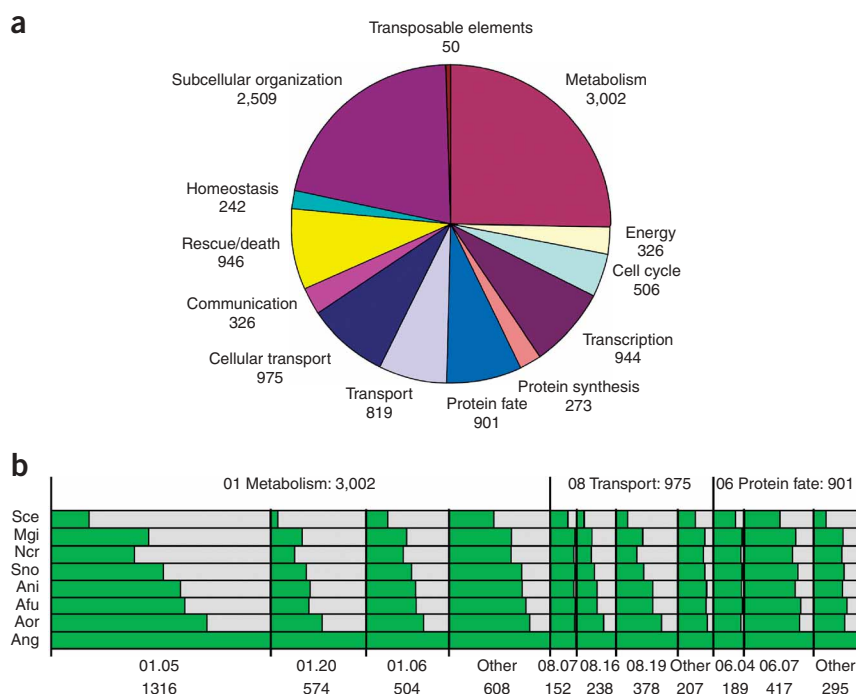


Figure 1 Functional classification of *A. niger* ORFs. **(a)** FunCat classes¹² are, 01, Metabolism; 02, Energy; 03, Cell cycle and DNA processing; 04, Transcription; 05, Protein synthesis; 06, Protein fate (folding, modification, destination), 07, Transport facilitation; 08, Cellular transport and transport mechanisms; 10, Cellular communication/signal transduction mechanism; 11, Cell rescue, defense and virulence; 13, Regulation of/interaction with cellular environment; 40, Subcellular localization; and 29, Transposable elements. The numbers following class names represent the number of genes in each category. **(b)** Comparison of functional classes in different fungi. Sce, *Saccharomyces cerevisiae*; Mgi, *Magnaporthe grisea*; Ncr, *Neurospora crassa*; Sno, *Stagonospora nodulum*; Ani, *Aspergillus nidulans*; Afu, *Aspergillus fumigatus*; Aor, *Aspergillus oryzae*; Ang, *Aspergillus niger*. Metabolism (01): C-compound and carbohydrate metabolism (01.05), lipid, fatty-acid and isoprenoid metabolism (01.06) and secondary metabolism (01.20); Cellular transport (08): vesicular transport (08.07), extracellular transport, exocytosis and secretion (08.16) and cellular import (08.19); Protein fate (06): protein secretion and fate (06.04) and cellular transport (06.07). The lengths of the bars are proportional to the number of genes in each category.

A. nidulans centromere-flanking genes mapped without exception to either the 5' or 3' end of a large *A. niger* supercontig and for each linkage group two centromere flanking supercontig ends were found. However, no clear centromeric sequences were identified. Inferred telomeric regions show little to no synteny (Fig. 2).

Linking the *A. niger* physical map to the genetic map

Parasexual recombination and electrophoretic karyotyping were previously used to establish the presence of eight linkage groups in the widely studied *A. niger* strain CBS 120.49 (refs. 14,15). We used these methods to assign 72 characterized or newly selected genes to specific chromosomes (Supplementary Table 4 online). This positioning unambiguously links all supercontigs of CBS 513.88 to single chromosomes of CBS 120.49. Electrophoretic karyotype data indicate that chromosomes VI and VII may not be completely covered by sequence data (Table 2). For chromosome VI this most likely relates to the presence of rDNA repeats whereas for chromosome VII this may be due to a large deletion in strain CBS 513.88. Four different methods were used to establish the orientation of supercontigs on the chromosomes resulting in a proposed alignment of the 19 supercontigs of strain CBS 513.88 with the eight linkage groups of strain CBS 120.49 (Table 2 and Supplementary Data online).

Life cycle and reproduction

The majority of aspergilli, including *A. niger* are only known to reproduce by asexual means, forming conidiospores¹⁶. The identified set of genes involved in signal transduction and conidiophore development is essentially the same as that of *A. nidulans* (Supplementary Table 5 online). Asexuality is thought to be derived from an ancestral sexual state¹⁷. We therefore screened the genome of *A. niger* for the presence of genes involved in incompatibility (Supplementary Table 6 online) and sexual reproduction (mating processes, signal transduction and ascumata development). A full complement of apparently

functional, early sexual development genes, including a key *mat-1* alpha domain mating-type gene was identified (Supplementary Table 5 and Supplementary Data). Two genes were only partially retrieved and one (*pro1*) contained a premature stop codon. *mat-1* and pheromone-precursor and receptor genes were not expressed during fed-batch fermentation, but this was not surprising given that mating and pheromone signaling require specific environmental conditions for induction (Supplementary Table 7 online). It has recently been suggested that the 'asexual' species, *A. fumigatus* and *A. oryzae* might have sexual potential^{9,18}. A similar situation could apply to *A. niger*, which would be of great value for strain improvement.

Cell wall development

The fungal cell wall determines biotechnologically relevant features such as morphology during fermentation and cell integrity. The cell wall of *A. niger* consists of chitin, 1,3- β -glucan, 1,6- β -glucan, 1,3- α -glucan, galactosaminogalactan and galactomannan, similar to *A. fumigatus*¹⁹. Genes required for the biosynthesis of these cell wall components were identified in the genome (Supplementary Table 8 online). As in other aspergilli, a high degree of redundancy is observed for chitin synthases and chitinases compared to *Saccharomyces cerevisiae* (Supplementary Table 9 online). *A. niger* contains five putative 1,3- α -glucan synthase genes, the highest number in the four published *Aspergillus* genomes. Specific members of each gene family were expressed during vegetative growth in a fed-batch culture on glucose and ammonium (Supplementary Table 7).

The cell wall integrity (CWI) signaling pathway is responsible for cell wall remodeling and reinforcement in response to a changing environment²⁰. Several components of this pathway are present in the genome (Supplementary Table 8) and indeed expressed during fed-batch culture, ensuring the growth of *A. niger* under harsh industrial-process conditions (Supplementary Table 7 and Supplementary Fig. 2 online).

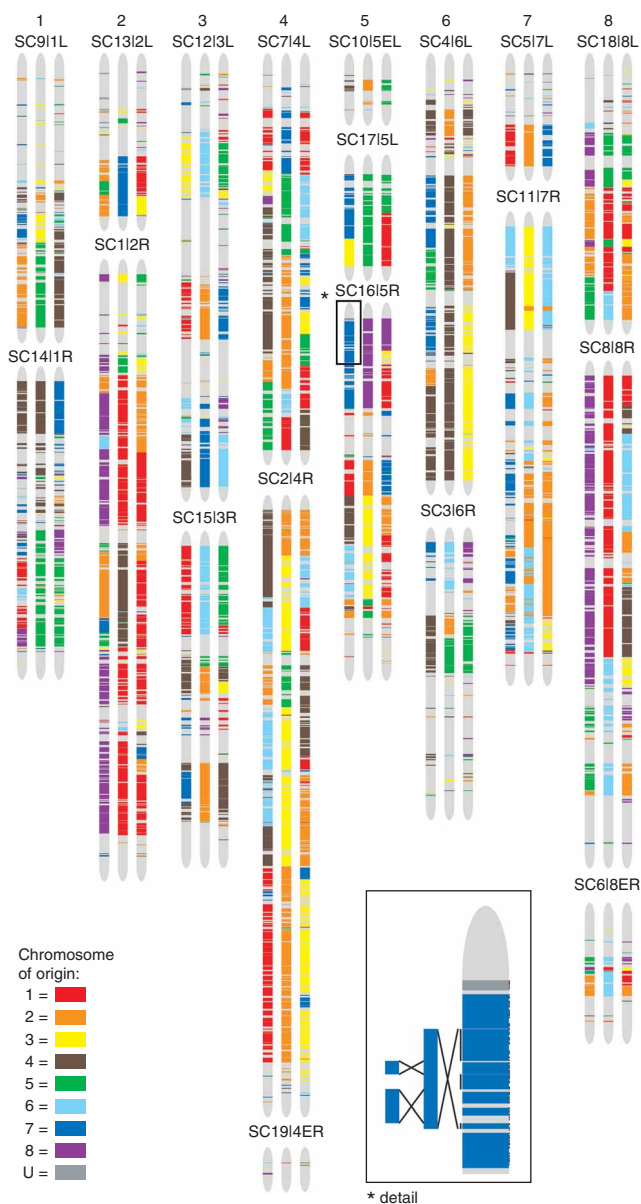


Figure 2 *Aspergillus* comparative genome maps. *A. niger* supercontigs (SC1-SC19) are ordered from left to right by their proposed chromosomal assignment (1 to 8) and are projected vertically, left arm (L) on top. Syntenic regions between *A. niger* and *A. nidulans*, *A. fumigatus* and *A. oryzae* are plotted in three separate columns on the *A. niger* chromosomes. The colored areas refer to the chromosome assignment in the compared aspergilli (linkage group of origin). Gray blocks are syntenic with contigs for which this chromosome assignment is unknown (U). The zoom-in on SC16 displays the position of orthologs on the chromosome (black bars at the right) and regions of reverse orientation (black bars at the left).

Hydrophobins are important determinants of cell morphology through their amphipathic monolayer structure²¹ (**Supplementary Table 8**). Two out of seven of the predicted hydrophobin genes are expressed during submerged fermentation (**Supplementary Table 7**). By modifying the expression of hydrophobin genes, the cell wall composition and hyphal morphology of *A. niger* may be altered, for example, to improve protein production²². Several genes involved in autolysis are present^{16,23}. Reduced autolysis through controlled expression of these genes is relevant for industrial processes to minimize proteolysis of secreted proteins and to simplify downstream processing²⁴.

Central metabolism and organic acid production

A novel genome-scale model of *A. niger* central metabolism was developed on the basis of the *A. niger* genome. This metabolic model was constructed by supplementing literature on *A. niger* metabolism with pathway databases and literature on other aspergilli^{25,26}. The reconstructed metabolic network consists of 1,069

unique reactions, of which 733 are based directly on literature and 832 are supported by genomic data, with an overlap of 570 reactions. Reactions not supported by literature or genomic data were included to ensure connectivity of the different pathways. For central metabolism only 24 enzymes out of 785 (excluding transport) are not linked to an annotated open reading frame (ORF, 3%), indicating an excellent coverage of metabolic genes. An overview of the subset of 785 unique reactions is presented in **Supplementary Table 10** online. Examining the entire metabolic model, only 65 enzymes were not found in the annotation (**Supplementary Table 11** online). A more detailed discussion on *A. niger* central metabolism and transport characteristics is presented in **Supplementary Data, Supplementary Figures 3 and 4** online.

The extremely flexible metabolism and high nutritional versatility of *A. niger* is confirmed by the presence of various solute-transporter classes (**Supplementary Table 12** online). Members of the major facilitator superfamily appear to be exceptionally abundant in *A. niger* with 461 genes, which is in the same range as *A. oryzae* (507), but greater than in *A. fumigatus* (275) and *A. nidulans* (358) (**Supplementary Table 13** online). These proteins are predicted to be involved in the transport of a wide range of substrates and several may function as nutrient sensors. *Aspergillus* species have also undergone a significant expansion in the number of fungal zinc binuclear cluster proteins (PF00172)²⁷ (**Supplementary Table 14** online). This distribution is mirrored by a second PFAM domain (PF04082), which was originally identified as occurring in a proportion of proteins with PF00172 domain. This is most dramatic in *A. niger* and is not reflected in other transcription factor domains (**Supplementary Table 15** online). The characterized examples of these proteins are involved in regulating diverse aspects of primary and secondary metabolism including polysaccharide degradation.

Using the reconstructed metabolic network, we investigated why *A. niger* is an efficient organic acid producer²⁸. Several enzymes involved in the formation of the citrate precursor oxaloacetate were found, for example, two pyruvate carboxylases (one cytoplasmic, one mitochondrial), four malate dehydrogenases (three cytoplasmic and one mitochondrial). In addition a mitochondrial oxaloacetate transporter was found (**Fig. 3**). *A. niger* contains one cytosolic and three putative mitochondrial citrate synthases. Phylogenetic analysis of the *A. niger* citrate synthases in various aspergilli reveals two clusters containing one and three members, respectively (**Supplementary Fig. 5**). The second cluster may have originated from a duplication of an ancestral citrate synthase that was lost in all organisms except in some fungi and plants. At least one additional gene duplication event occurred (An01g09940 and An08g10920). Similar redundancy was found for aconitase, for which two putative cytoplasmic and two mitochondrial forms were found. The various gene duplications may be important in view of the efficient production of citrate by *A. niger*. Moreover, the necessary transport steps could be facilitated by two mitochondrial and several cytoplasmic membrane tricarboxylate

Table 2 *A. niger* genetic and physical map

Chromosome	Supercontig		Mapped genes linking chromosomes to contigs		Orientation	Orientation evidence		Supercontig size	Total size	Estimated chromosome size
			By CHEF hybridization	By parasexual analysis		Synteny expression	Mapped genes			
I	L	14	<i>eglA, argB, rhgB, pelA, pepA, aguA</i>	<i>argB, pepA</i>	← c	s, e	nid 1	1.7	3.3	3.5
	R	9	<i>faeA, abnA, rgaeA, pgaD, bphA, pepD</i>	<i>bphA, pepD</i>	c ←	s, e	nid 1	1.6		
II	L	13	<i>alcB</i>		← c	s, e		0.9	4.5	5.0
	R	1	<i>abfA, pepB, xyrA, kexB, xlnD, pgal, cbhB, pgaE</i>	<i>cpcA, pepB, goxC</i>	c →	s, e	nid 7	3.6		
III	L	12	<i>rhgA, pyrA, axeA, pgaX, glkA, areA, faeB</i>	<i>pyrA, areA</i>	← c	s, e	nig 1	2.6	4.5	4.1
	R	15	<i>abfB, pgalI, xlnR</i>	<i>xlnR</i>	c →	s, e		1.9		
IV	L	7	<i>xkiA, pepC, pepF, eglB, rodA, cbhA</i>	<i>xkiA</i>	→ c	s, e		2.3	6.0	6.6
	R	2	<i>pgaB, mpdA, pepE, pacC, aglB, hxA</i>	<i>creA, pkaC, mpdA, pepE, pacC</i>	c ←	s	nid 6	3.6		
	R	19	<i>pelD</i>		c ← ←	e		0.1		
V	L	10	<i>plyA</i>	<i>prtF</i>	→ → c	e		0.3	3.0	3.5
	L	17	<i>gatA</i>		→ c	s	nid 3	0.6		
	R	16	<i>gpdA, pgaA</i>	<i>pkaR</i>	c ←	s, e	nid 1	2.1		
VI	L	4		<i>glcA, acuA</i>	← c	t, s, e	nid 3	2.5	4.1	5.2
	R	3	<i>pelB, axhA, rdna, pmeA, glaA</i>		c ←	s, e	nid 2	1.6		
VII	L	5	<i>pgaC</i>		← c	s, e		0.6	3.3	4.1
	R	11	<i>pelC, bipA, nicB</i>	<i>nicB</i>	c ←	s, e	nig 1	2.7		
VIII	L	18	<i>nirA, bglA</i>	<i>nirA</i>	→ c	s, e	nig 1	1.5	5.1	5.0
	R	8	<i>cprA</i>	<i>trpC, acuB, niaD</i>	c →	s	nig 1	2.9		
	R	6	<i>aglA</i>		c → →	e		0.7		

The eight linkage groups were numbered in the order of which new markers were characterized and were later shown to correspond to chromosomes I to VIII¹⁵. The chromosome numbers do not reflect their size. The proposed position of the supercontigs on the left arm (L) or right arm (R) of the chromosomes (I–VIII) is indicated. The arrows indicate the orientation of the supercontigs relative to the centromere (c). Orientation evidence is based on telomeric sequences (t), synteny with *A. nidulans* near centromere (s), expression data showing telomeric positioning effect (e) and genetically mapped and identified *A. niger* genes (nig). The nig value indicates the number of mapped *A. niger* genes. The nid value is the number of orthologous *A. nidulans* genes corresponding to *A. niger* markers that confirm the proposed orientation. Total size refers to the summed supercontig sizes whereas the chromosome size is based on previous predictions¹⁵. All sizes are in Mbp.

transporters. *A. niger*, like other fungi, contains two ATP-citrate lyases and one eukaryotic beta-chain mitochondrial citrate lyase. This indicates a high degree of conservation of the 'fermentative pathway' of citrate. Because one of the ATP-citrate lyases is mitochondrial, a futile cycle of citrate formation and degradation needs to be prevented. The identification of genes involved in citric acid metabolism and transport provides excellent opportunities to study and understand the efficient production of citric acid by *A. niger* in much more detail.

A. niger is also able to efficiently produce gluconic acid. Genes are present that encode one intracellular and three secreted glucose oxidases. The presence of eleven ORFs encoding catalases ensures protection against hydrogen peroxide generated by glucose oxidase;

two ORFs contain a signal sequence required for protein export. At least one of the four putative lactonases is located extracellularly. Two gluconate-specific kinases are present, suggesting that catabolism of gluconate proceeds via phosphorylation to 6-P-gluconate.

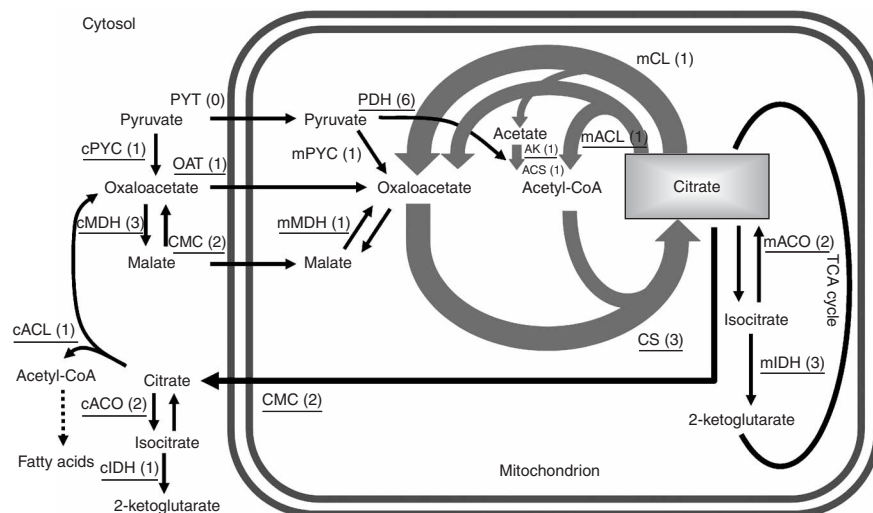
Oxalic acid is an undesired by-product in *A. niger* fermentations. One biosynthetic pathway involves oxaloacetate hydrolase²⁹. In addition, the presence of two glycolate reductases and several lactate dehydrogenases in *A. niger* suggests the activity of an alternative pathway. This resembles the situation in mammalian liver cells³⁰, in which the glyoxylate cycle operates during citrate consumption. Oxalate may be further catabolized by oxalate decarboxylase, yielding formate. Four putative oxalate/formate antiporters for formate transport were found (Supplementary Table 12).

Figure 3 Citrate biosynthesis and degradation.

The putative localization of enzymes and transporters is indicated. Their symbols are underlined when expression of their corresponding *A. niger* genes (under noncitrific acid-producing conditions) was observed; the numbers are in bold in this legend. The numbers in parentheses next to the enzyme symbols indicate the number of genes found for a specific function whereas in the case of the pyruvate dehydrogenase complex they indicate the total number of genes involved in the complex.

cPYC, cytoplasmic pyruvate carboxylase (An04g02090); cMDH, cytoplasmic malate dehydrogenase (An11g07190, An12g00160, An15g00070); cACO, cytoplasmic aconitase (An02g11040, An09g03870); cIDH, NADP-specific isocitrate dehydrogenase (An02g12430; note that cytoplasmic or mitochondrial localization occurs by differential translational start point utilization); cACL, cytoplasmic ATP-citrate lyase (An11g00530);

PYT, putative pyruvate carrier, not identified because of lack of characterization in other organisms; OAT, oxaloacetate transporter (An14g06860); CMC, citrate/malate carrier (An11g11230, An18g00070); PDH, pyruvate dehydrogenase complex E1- α subunit (An07g09530, An11g04550), E1- β subunit (An01g00100), E2 dihydrolipoamide acetyltransferase (An07g02180), E3 lipoamide dehydrogenase (An07g06840), E3 binding protein (An02g11910); mPYC, mitochondrial pyruvate carboxylase (An15g02820); mMDH, mitochondrial malate dehydrogenase (An16g00120); mCL, mitochondrial citrate lyase (An01g08610); AK, acetate kinase (An02g06420) and ACS, acetyl-CoA synthase (An05g00390); CS, citrate synthase (An01g09940, An08g10920, An09g06680); mACL, mitochondrial ATP-citrate lyase (An11g00510); mACO, mitochondrial aconitase (An08g10530, An16g05760); mIDH, mitochondrial NADP- and NAD-isocitrate dehydrogenases (An02g12430, An08g05580, An18g06760).



Protein secretion in *A. niger*

For reasons unknown, *A. niger* is a far more effective natural secretor of proteins than the well-studied yeast *S. cerevisiae*³¹. In eukaryotes protein secretion involves transport via endoplasmic reticulum (ER), Golgi apparatus and vesicles to the cell membrane (Fig. 4 and Supplementary Table 16 online). Translocation from the cytoplasm to the ER occurs in *A. niger* through the established signal recognition particle (SRP)-dependent and SRP-independent pathways. As in mammals, no ortholog of the essential yeast signal recognition and docking protein Srp21p was found. ER luminal HSP70-type protein-folding chaperones, including BipA and LhsA are present. In *S. cerevisiae*, Kar2p (equivalent of BipA in *A. niger*) functions together with the nucleotide-exchange factor Sil1p³². Surprisingly, no ortholog of yeast Sil1p can be

found in *A. niger* and other aspergilli, whereas a putative ortholog in *Neurospora crassa* could be detected. *A. niger* encodes three soluble luminal and one putative membrane-bound protein disulphide isomerase (PDI). Only PdiA (ortholog of yeast Pdi1p) and the membrane-associated EpsA (Eps1p) appear to be close orthologs of known yeast proteins. Finally, in contrast to several yeast species, plants and mammals, the ER luminal protein EroA of the sequenced aspergilli contains a predicted C-terminal ER-retention signal.

The *A. niger* unfolded protein response (UPR) signaling pathway is strikingly different from that in yeast^{31,33}. In mammalian cells the protein p58 interacts with and inhibits the ER-localized eIF2 α kinase PERK. This kinase is involved in translational regulation, attenuating the UPR during ER stress³⁴. Orthologs of p58 and PERK are not found in yeast. The presence of a putative p58 ortholog in *A. niger* is remarkable as no ortholog of PERK was found.

Most components of the glycosylation machinery are readily identified by sequence similarity. Although present in *A. fumigatus* and *A. nidulans*, no ortholog of yeast Alg14p was identified in *A. niger*. Interestingly, the dolichol-P-mannose synthase gene is more mammalian- than yeast-like as it does not contain a C-terminal hydrophobic transmembrane region. *A. niger* possesses a glycosylation-dependent quality-control system that is clearly distinct from that of *S. cerevisiae*. *A. niger* contains an ortholog of the protein-folding sensor UDP-Glc:glycoprotein glucosyltransferase absent in yeast. Calnexin is more similar to its counterpart in *Schizosaccharomyces pombe*.

ER-associated protein degradation (ERAD) directs misfolded or unassembled proteins to the proteasome³⁵. *A. niger* lacks clear orthologs of the yeast ERAD proteins Cue1p, Rad23p, Ubx2p and Yos9p and shows little homology to other ERAD components such as Der1p, Hrd1p, Doa10p and Hrd3p. Orthologs of all subunits of the yeast 26S proteasome are found, but orthologs of the regulatory proteins Rpn13p and Rpn14p³⁶ appear to have low similarity in *A. niger* or to be absent (Supplementary Data). The *A. niger* secretion system is equipped with a well-developed machinery for glycoprotein

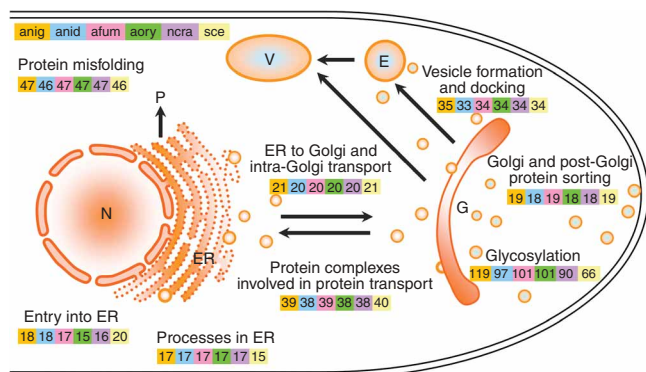


Figure 4 The secretion pathway of *A. niger* and other fungi. The number of identified genes from *A. niger* (anig) and their putative orthologs in several other fungi (anid, *A. nidulans*; afum, *A. fumigatus*; aory, *A. oryzae*; ncr, *N. crassa* and sce, *S. cerevisiae*) are shown for various steps of the secretory pathway. E, endosome; ER, endoplasmic reticulum; G, Golgi; P, proteasome; V, vacuole; N, nucleus. See Supplementary Table 16 for details.

Table 3 Comparison of carbohydrate active enzymes in aspergilli

Enzyme	CAZy family	<i>A. niger</i>	<i>A. nidulans</i>	<i>A. fumigatus</i>	<i>A. oryzae</i>
Glycosyl hydrolases					
β-glucosidase	GH1	3	3	5	3
β-galactosidase	GH2	0	3	2	1
β-mannosidase	GH2	3	3	2	3
β-glucuronidase	GH2	3	3	1	2
Exo-β-d-glucosaminidase	GH2	0	0	2	1
β-glucosidase	GH3	12	15	12	16
β-xylosidase	GH3	2	3	2	5
β-N-acetylhexosaminidase	GH3	3	2	3	2
β-1,4-endoglucanase	GH5	4	4	6	4
β-1,4-endomannanase	GH5	1	6	2	2
Exo-1,3-β-glucanase	GH5	4	3	2	3
Endo-1,6-β-glucosidase	GH5	0	1	1	2
Endoglycoceramidase	GH5	1	1	2	2
Cellobiohydrolase (non-reducing end)	GH6	2	2	1	1
Cellobiohydrolase (reducing end)	GH7	2	2	2	2
β-1,4-endoglucanase	GH7	0	1	2	1
β-1,4-endoxylanase	GH10	1	3	4	4
β-1,4-endoxylanase	GH11	4	2	3	4
Xyloglucan-active β-1,4-endoglucanase	GH12	3	1	4	4
α-amylase	GH13	4	3	2	3
α-1,6-glucosidase	GH13	2	2	5	5
α-1,3-glucan synthase	GH13	5	2	3	3
GPI-anchored α-amylase-like	GH13	3	3	3	2
Intracellular α-amylase-like	GH13	2	1	1	2
Glycogen debranching enzyme	GH13	2	2	1	2
Glucoamylase	GH15	2	2	5	3
Endomannanase	GH26	1	3	0	1
α-galactosidase	GH27	5	3	5	3
Endopolygalacturonase	GH28	7	3	4	4
Endorhamnogalacturonase	GH28	6	1	2	6
Exopolygalacturonase	GH28	4	3	3	2
Exorhamnogalacturonase	GH28	3	1	2	4
Xylogalacturonan hydrolase	GH28	1	1	1	2
α-l-fucosidase	GH29	1	0	0	0
α-glucosidase	GH31	7	9	6	10
α-xylosidase	GH31	0	1	1	1
Endoinulinase	GH32	1	0	1	0
Exoinulinase	GH32	1	1	1	0
Invertase	GH32	3	1	2	4
β-galactosidase	GH35	5	4	5	6
α-galactosidase	GH36	2	4	4	3
α-l-arabinofuranosidase	GH43	2	3	3	3
Endoarabinanase	GH43	4	4	6	6
β-xylosidase	GH43	2	5	2	5
β-1,4-xylosidase/α-arabinofuranosidase	GH43	1	1	4	3
Galactan 1,3-β-galactosidase	GH43	1	2	1	3
β-1,4-endoglucanase	GH45	0	1	1	0
α-l-arabinofuranosidase	GH51	3	2	2	3
Endo-1,4-β-galactanase	GH53	2	1	1	1
α-l-arabinofuranosidase	GH54	1	1	1	1
β-1,4-endoglucanase	GH61	7	9	7	8
Arabinoxylan arabinofuranohydrolase	GH62	1	2	2	2
α-glucuronidase	GH67	1	1	1	1
Endoglucanase	GH74	1	2	2	0
α-l-rhamnosidase	GH78	8	8	6	8
d-4,5 unsaturated glucuronyl hydrolase	GH88	1	2	2	3
Exoarabinanase	GH93	0	2	3	3
α-l-fucosidase	GH95	2	2	2	3

Continued on next page.

Table 3 continued

Enzyme	CAZy family	<i>A. niger</i>	<i>A. nidulans</i>	<i>A. fumigatus</i>	<i>A. oryzae</i>
Rhamnogalacturonyl hydrolase	GH105	2	3	2	2
Polysaccharide lyases					
Pectate lyase	PL1	1	3	2	4
Pectin lyase	PL1	5	5	4	8
Pectate lyase	PL3	0	5	3	3
Rhamnogalacturonan lyase	PL4	2	4	3	4
Pectate lyase	PL9	0	1	1	1
Rhamnogalacturonan lyase	PL11	0	1	0	0
Carbohydrate esterases					
Acetyl xylan esterase	CE1	1	1	2	1
Feruloyl esterase	CE1	1	1	0	2
Unknown esterase	CE1	1	1	2	2
Pectin methylesterase	CE8	3	3	4	5
Rhamnogalacturonan acetyl esterase	CE12	2	2	2	4
Total		171	186	186	217
Secreted		131 (77%)	134 (72%)	124 (66%)	159 (73%)
Intracellular		40 (23%)	52 (28%)	62 (34%)	58 (27%)

ORFs are grouped by CAZy family and putative function based on the closest characterized homolog. GPI, glycosyl phosphatidylinositol.

quality-control, combined with an effective ERAD pathway, which differs from that in yeast. In conclusion, the *A. niger* secretory system shares some components with mammals and some with *S. cerevisiae*.

Biopolymer-degrading hydrolases

Aspergilli contain a wide spectrum of enzymes for polysaccharide, protein and lipid degradation. The two industrial aspergilli (*A. niger* and *A. oryzae*) contain the highest percentage of extracellular enzymes. Cellulases, hemicellulases, pectinases, amylases, inulinases, lipases and proteases are used in a range of industrial applications. Here we focus on polysaccharide-degrading enzymes and proteases.

Glycosyl hydrolase, lyase and esterase families involved in polysaccharide degradation in the aspergilli sequenced were identified using the carbohydrate-active enzymes (CAZy) classification (<http://www.cazy.org/>, Table 3). A detailed overview of the *A. niger* enzymes is presented in Supplementary Table 17 online. Specific differences between the aspergilli were noticed. In contrast to the other three aspergilli, *A. niger* contains only one GH10 and four GH11 endoxylanases, a xylanase family liberating larger oligosaccharides. For inulin degradation *A. niger* and *A. fumigatus* both contain invertases and endo- and exoinulinases. However, *A. nidulans* does not have endoinulinases, whereas *A. oryzae* appears to contain only invertases. This suggests that the latter two species degrade inulin extracellularly.

For the degradation of pectin a large set of enzymes is required³⁷. *A. niger* has the largest set (21) of (GH28) enzymes, including seven candidate endopolygalacturonases (Supplementary Table 17). In contrast to the other aspergilli, *A. niger* has only a single pectate lyase (PL1). The pectinolytic machinery is completed by a single putative delta-4,5-unsaturated glucuronyl hydrolase (GH88), two putative delta-4,5-unsaturated rhamnogalacturonyl hydrolases (GH105), three putative pectin methylesterases (CE8) and one candidate rhamnogalacturonan acetyl esterase (CE12) (Table 3). The single pectate lyase is consistent with the acidifying properties of *A. niger*, given that pectate lyases show little activity at acidic pH. Moreover, *A. niger* does not contain exoarabinanases, indicating that *A. niger*

relies on the combined action of endoarabinanase and α -L-arabinofuranosidase for the hydrolysis of the arabinan side-chains of pectin.

The glycoside hydrolase family GH13 involved in starch degradation is larger in aspergilli compared to other fungi. The GH13 family contains three separate groups of amylase-type enzymes in *A. niger* based on phylogenetic clustering (Supplementary Fig. 6). *A. niger* contains four putative extracellular α -amylases (Supplementary Table 17). A second group consists of enzymes with relatively low homology to known alpha-amylases, which do not contain signal sequences. A third group consists of glycosyl phosphatidylinositol-anchored enzymes, which may play a role in the maintenance of the cell walls containing α -glucan³⁸. The *A. niger* genome contains 42 enzymes classified as CAZy family members without a signal sequence. The majority of these are likely to be involved in the degradation of disaccharides or glycosides imported by the fungus.

A. niger has a full set of enzymes to degrade polypeptides (Supplementary Table 18 online). Secreted proteases are applied in detergents, food applications and as biocatalysts in the production of fine chemicals. The *A. niger* genome encodes 198 proteins involved in proteolytic degradation including a variety of secreted aspartyl endoproteases (9), serine carboxypeptidases (10) and di- and tripeptidylaminopeptidases (9). Compared to *A. nidulans* and *A. oryzae*, the number of putative secreted aminopeptidases (3) is low. The abundance of aspartyl endoproteases and carboxypeptidases (mostly active at low pH) and the low number of aminopeptidases (mostly active at neutral or high pH) matches the acidifying properties of *A. niger* as previously noted for *A. oryzae*⁸. Unlike the aminopeptidases in the other sequenced aspergilli, seven aminopeptidases of *A. niger* are predicted to be intracellular, indicating that at least part of the amino-terminal degradation of external proteins is taken care of by secreted di- and tri-peptidylpeptidases.

Secondary metabolism and safety

Among the secondary metabolites produced by filamentous fungi, mycotoxins are most relevant from a safety point of view³⁹. *A. niger*

contains several secondary metabolite clusters (**Supplementary Table 19** online). The genome contains 17 nonribosomal peptide synthase (NRPS) and 34 polyketide synthase (PKS)-encoding genes, most of which are located in clusters. *A. niger* contains seven hybrid PKS-NRPS synthases whereas other sequenced filamentous fungi contain a single hybrid PKS-NRPS. The sometimes unusual domain patterns for these NRPS, PKS and hybrid genes are given in **Supplementary Table 20** online. Of particular interest is the presence of a gene cluster that shares a number of genes with the *Gibberella moniliformis* gene cluster, which encodes the mycotoxin fumonisin (**Supplementary Fig. 7**)⁴⁰. This cluster is absent from the genomes of other sequenced aspergilli such as *A. fumigatus*, *A. oryzae* and *A. nidulans*^{8–10}.

A putative ochratoxin cluster was identified on the basis of a PKS fragment of *A. ochraceus* involved in ochratoxin biosynthesis⁴¹. Some *A. niger* strains have been reported to synthesize this toxin but little is known about the biosynthetic pathway⁴². In *A. ochraceus*, genes encoding chloroperoxidase, reductase, esterase, dehydratase and NRPS may be involved. Of these only an NRPS-like gene with the unusual domain structure ACPA is found in *A. niger*. Neither gene is expressed at a detectable level (**Supplementary Table 7**).

Siderophores are secondary metabolites involved in iron assimilation and storage⁴³ and protection against oxidative stress. One clear example of a siderophore cluster was identified in the genome (**Supplementary Table 19**). The vast majority of the remaining nonribosomal peptide and polyketide synthases do not share orthologs with the other fungal genomes sequenced to date. In addition, no orthologs of genes for the biosynthesis of penicillin, prenylated alkaloids (fumitremorgins), clavines, gliotoxin, aflatoxin or known fungal terpenes (aristolochene and trichodiene) were found.

METHODS

Genome sequencing and assembly. A library with $\sim 20.7\times$ coverage of the *A. niger* CBS 513.88 genome was constructed in pBeloBAC 11 and characterized by end-sequencing and restriction digestion. Insert sizes of BAC clones ranged from <50 to >150 kb per clone. A total of 11,136 BAC clones were generated with an average insert size of 68 kb. We generated 18,014 BAC end-sequences with 504 bases average read length (phred20). The BAC end-sequencing success rate was 80.9%. Using the BAC walking approach, we selected 595 BAC clones for analysis, and 505 of those, representing the minimal tiling path, for shotgun sequencing. The minimal tiling set of sequenced BACs was confirmed by the high density map of BAC end-sequences (1 end per 2 kb). Repeated sequencing of the estimated 100 rDNA copies was avoided by not sequencing BACs that have end sequences matching rDNA. The 505 BACs selected have an average insert size of 76.8 kb and have been sequenced with a $7.5\times$ coverage. The assembled BACs contained on average 2.14 contigs per BAC, corresponding with an average of 1.14 sequence gaps per BAC. Shotgun sequences were assembled with Phrap and edited after import into Gap4. BAC assemblies and raw data were visualized and edited using the STADEN package. The genome was assembled and logically joined using BAC clones physically bridging known gaps to form supercontigs with a total unique sequence size of 33.9 Mb. The estimated 5% of the genome not sequenced includes telomeric regions, additional rDNA repeats and small gaps. Full details of the genome sequencing and assembly can be found in **Supplementary Methods** online. CBS 513.88 can be obtained from Centraal Bureau Schimmelcultures (<http://www.cbs.knaw.nl/>).

Annotation. Analysis and annotation of the genomic sequences of *A. niger* was performed with a combined automatic and manual approach. ORFs were predicted by a version of FGENESH⁴⁴ trained on known *A. niger* and related organism sequences as well as other gene prediction algorithms (**Supplementary Methods**). ORFs were named after the organism (An), supercontig number (two digits) followed by g (gene) and a five-digit number matching the order of the ORFs on the contig. tRNA genes were identified as described in **Supplementary Methods**. For all ORFs identified, exhaustive automatic

bioinformatic analysis with respect to function and structure of the respective protein was carried out using the PEDANT-ProTM software⁴⁵. For each ORF the automatically predicted functional features were manually verified and the following features were manually annotated: gene model (based on Blast comparisons with known proteins), title, functional categories according to the MIPS functional catalog¹² and EC numbers.

The manual annotation of the title involved (i) selection of a similar protein through Blast comparison and (ii) a description of the degree of similarity with this protein. If possible a similar protein with published experimental data was selected. The description of the degree of similarity was assigned as follows: questionable ORF, no Blast hit and questionable gene structure; no similarity, no Blast hit; weak similarity, Blast e-value $>e-05$; similarity, Blast e-value $<e-05$ and $>e-20$; strong similarity, Blast e-value $<e-20$; known protein, identical to known *A. niger* gene.

Manual FunCat and EC number assignments were based on the described function of the selected similar protein and took into account the degree of similarity as illustrated by the following example. An *A. niger* protein matching a well-known transcriptional activator was annotated in the manual FunCat as follows: weak similarity, 04.05 mRNA transcription; similarity, 04.05.01 mRNA synthesis; strong similarity, 04.05.01.04 transcriptional control; known protein, 04.05.01.04.01 transcriptional activator.

Phylogenetic analysis. Orthologous genes were grouped using all-against-all, pairwise Blast similarity searches at the level of predicted proteins keeping reciprocally best-matching genes. Twenty fungal panorthologous genes encoding housekeeping functions and showing reciprocal best matches between all selected organisms were selected (**Supplementary Table 3**) and the predicted protein sequences were aligned with ClustalW⁴⁶. Bidirectional Blast similarity searching using the Blastx and tBlastn algorithms was applied to find additional orthologs in the genome of *A. clavatus*. After deletion of ambiguously aligned regions, the protein sequences were concatenated (7,767 amino acid sites). Maximum likelihood phylogenetic analysis was performed with Tree Puzzle⁴⁷ using the VT model⁴⁸ and a gamma model of rate heterogeneity with $\alpha = 0.65$. Reliable bootstrap values were obtained for all nodes of the tree except for the terminal nodes linking *A. niger*, *A. oryzae* and *A. terreus*, indicating their close relationship.

For phylogenetic analysis of protein families, the protein sequences were aligned first with ClustalW or ClustalX. Phylogenetic analyses were carried out in MEGA 2.1 or 3.1 using Maximum Likelihood, Neighbor Joining and Minimum Evolution⁴⁹. Stability of clades was evaluated by 500 to 1,000 bootstrap rearrangements. Genome sequences used were obtained from the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/Genomes/>) except for *A. clavatus* (http://msc.tigr.org/aspergillus/aspergillus_clavatus_nrrl_1/index.shtml) and *A. terreus* (http://www.broad.mit.edu/annotation/genome/aspergillus_terreus/).

Ortholog detection and synteny analysis. To detect conservation of gene order, we obtained pairwise lists of orthologs using bidirectional Blast searching of the *A. niger* predicted coding sequence (CDS) and predicted CDS of each of *A. fumigatus*, *A. nidulans* and *A. oryzae*. Sequences having a length ≤ 100 amino acids were omitted to get high confidence e-values. Bidirectional best hits having an e-value $<1e-10$ in both directions were considered as orthologous gene pairs. Conservation of synteny was determined by comparing adjacent orthologous gene pairs allowing for inversions in gene order between species. If there was a match, the two orthologous gene pairs were considered to be part of a syntenous region. Such regions were extended by directional scanning along the *A. niger* genome. Gene order and inversions were also recorded.

***A. niger* genetic and physical map.** The genetic location of cloned genes was established using pulsed field electrophoresis followed by Southern blot analysis¹⁵. Parasexual analysis¹⁴ was used to determine the genetic location of *cpcA* by linkage of the phleomycin resistance to *hisD4*. The methods used to establish the physical map including supercontig orientation on chromosomes are described in the **Supplementary Data**.

Fermentation and transcriptional profiling. *A. niger* was grown on defined medium with glucose as a carbon source and ammonia as nitrogen source using 20-liter submerged stirred (Rushton turbines) fermenters at controlled

pH of 4.5 at 35 °C. Glucose feeding started after 24 h. Biomass samples for mRNA analysis were taken after 72 and 120 h. Sample treatment and Affymetrix array (DSM proprietary GeneChips) analysis is described in **Supplementary Methods**.

Reconstruction and analysis of the metabolic network. A comprehensive literature search was carried out to identify reactions and/or enzymes that are present in *A. niger*, and this compilation was supplemented by information present in the enzyme databases BRENDA (<http://www.brenda.uni-koeln.de/>) and SwissProt (<http://www.expasy.org/>). The KEGG database (<http://www.genome.ad.jp/kegg/metabolism.html>) and the reconstructed metabolic network of *S. cerevisiae*²⁶ were used to fill gaps in incomplete pathways. Finally, the information obtained through the annotation of all ORFs (see above) was compared to this metabolic framework and as many ORFs as possible were assigned to reactions. PSORT II (<http://psort.ims.u-tokyo.ac.jp/form2.html>) and TargetP (<http://www.cbs.dtu.dk/services/TargetP/>) were used for analysis of theoretical subcellular localization. Prediction of signal peptide cleavage sites was carried out with SignalP 2.0 (<http://www.cbs.dtu.dk/services/SignalP/>). Protein sequences were also analyzed with InterProScan (<http://www.ebi.ac.uk/InterProScan/>).

Carbohydrate-active enzyme classification. The search for carbohydrate-active enzymes was performed according to the routine update strategy of the CAZY database (<http://afmb.cnrs-mrs.fr/CAZY/>). Sequences of the proteins in CAZY were cut into their constitutive modules (catalytic modules, carbohydrate-binding module (CBMs) and other noncatalytic modules) and the resulting fragments were assembled in a sequence library for Blast searches⁵⁰. Each *A. niger* protein model was compared by Blast analysis against the library of around 50,000 individual modules. Models with an e-value <0.1 were manually analyzed to predict their function based on multiple sequence alignment using ClustalW and a search for conserved signatures/motifs characteristic of each family including the presence of the catalytic machinery. SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) was used to detect the presence of possible signal sequences.

Genome data and availability. The annotated genome has been deposited at the EMBL database with the accession numbers AM270980–AM270998 (DNA supercontigs) and AM269948–AM270415 (DNA contigs).

Requests for materials. A scalable vector graphic that plots all syntenous regions on the *A. niger* chromosomes, including relative orientation of the conserved syntenies and blocks representing each gene, is available on request from H.J.K. (harrie.kools@wur.nl).

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

E.G.J.D., P.M.C., B.H. and F.M.K. wish to acknowledge the financial support from the European Commission (STREP FungWall grant, contract: LSHB-CT-2004-511952) and the French Ministry of Research (program ACI-BCMS, Enzywall). The work of X.L., U.R., J.S. and A.-P.Z. was partly funded by the Sonderforschungsbereich 578 (SFB578) of the Deutsche Forschungsgemeinschaft, Germany. We acknowledge the Department of Energy, Joint Genome Institute, TIGR and the Broad Institute for allowing some comparative genome analysis. Part of this work was supported by SENTER (BTS project BTS00010, TSGE 3012). Array hybridizations were performed at the MicroArray Department (MAD) in Amsterdam. Ulrike Jacobi is acknowledged for assistance in transcriptome analysis. We thank Bea den Dekker for excellent organizational support.

AUTHOR CONTRIBUTIONS

H.J.P. coordinated the *A. niger* sequencing and annotation work. J.H.de W. contributed to the microarray analysis, and metabolism and developmental aspects. R.A. and K.A. were coordinators of the Biomax annotation team. M.v.d.B. and A.v.D. were members of the DSM sequencing team. J.L. and H.W. at Qiagen developed the sequencing strategy and coordinated the sequencing. N.N.M.E.v.P. was responsible for the microarray work. R.M. and S.B. were responsible for the fermentation. H.M. was involved in transcriptome analysis. G.S.P.G. contributed to the manual annotation. J.A.R. was responsible for the DSM bioinformatics work. M.O. contributed to protein analysis and characterization. P.J.S., R.T.J.M.v.d.H. and H.J.K. worked out the phylogeny and syntenies between the aspergilli. P.J.I.v.d.V. and A.J.M.De. compared the physical

and genetic map. P.S.D. and H.A.B.W. investigated life cycle and reproduction whereas K.P., J.Va. and A.J.M.De. looked for incompatibility genes. Cell wall functions were analyzed by A.F.J.R., F.M.K., P.W.J.de G. and H.A.B.W. Transcription factors were analyzed by M.X.C., A.M.M., S.G.O. and M.C. J.D.B. and D.U. analyzed the *A. niger* secretome. Analysis of the metabolism was carried out by G.H. (coordination), M.R.A., J.N., C.d'E., C.P.K., M.S. and J.Vi., whereas A.J.M.Dr. and P.A.v.K. analysed the genome for transport functions. D.B.A. coordinated the secretion team, which further included T.G., C.A.M.J.J.v.d.H., X.L., N.M.E.v.P., A.F.J.R., U.R., C.M.J.S., J.S. and A.-P.Z. The glycosylation aspects were covered by R.C., S.G. and W.V. The general analysis of the extracellular enzyme spectrum was performed by R.P.deV. (coordination), P.M.C., E.G.J.D., M.H. and B.H., whereas J.A.E.B. (pectinases), P.D. (proteases), R.M.v.d.K., M.J.E.C.v.d.M., C.G. and L.D. (amylases) covered specific areas. G.T., M.C. and P.W.M.v.D. analysed secondary metabolism and safety aspects. J.P.T.W.v.d.H. and A.J.J.v.O. initiated the DSM sequencing project. H.J.P., J.H.d.W., J.Vi. and H.S. (project leader) wrote the final text of the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at <http://www.nature.com/naturebiotechnology/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

- Pandey, A., Selvakumar, P., Soccol, C.R. & Nigam, P. Solid state fermentation for the production of industrial enzymes. *Curr. Sci.* **77**, 149–162 (1999).
- Berka, R.M., Dunn-Coleman, N. & Ward, M. Industrial enzymes from *Aspergillus* species. in *Aspergillus Biology and Industrial Applications* (eds. Bennett, J.W. & Klich, M.A.) 155–202, (Butterworth-Heinemann, Stoneham, MA, 1992).
- Schuster, E., Dunn-Coleman, N., Frisvad, J.C. & van Dijk, P.W. On the safety of *Aspergillus niger*—a review. *Appl. Microbiol. Biotechnol.* **59**, 426–435 (2002).
- Punt, P.J. et al. Filamentous fungi as cell factories for heterologous protein production. *Trends Biotechnol.* **20**, 200–206 (2002).
- Archer, D.B. & Turner, G. Genomics of protein secretion and hyphal growth in *Aspergillus*. in *The Mycota XIII* (ed. Brown, A.J.P.) 75–96, (Springer, Berlin Heidelberg, 2006).
- Ruijter, G.J.G., Kubicek, C.P. & Visser, J. Production of organic acids by fungi. in *The Mycota X Industrial Applications* (ed. Osiewacz, H.D.) 213–230, (Springer, Berlin Heidelberg, 2002).
- Raj, H.G., Saxena, M. & Allameh, A. Metabolism of foreign compounds by fungi. in *Handbook of Applied Mycology vol. 4: Fungal Biotechnology* (eds. Arora, D.K., Elander, R.P. & Mukerji, K.G.) 881–904, (Marcel Dekker Inc., New York, 1992).
- Machida, M. et al. Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* **438**, 1157–1161 (2005).
- Galagan, J.E. et al. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**, 1105–1115 (2005).
- Nierman, W.C. et al. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* **438**, 1151–1156 (2005).
- Nakamura, Y., Gojobori, T.S. & Ikemura, T. Codon usage tabulated from the international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**, 292 (2000).
- Ruepp, A. et al. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res.* **32**, 5539–5545 (2004).
- Samson, R.A. Current taxonomic schemes of the genus *Aspergillus* and its teleomorphs. in *Aspergillus Biology and Industrial Applications* (eds. Bennett, J.W. & Klich, M.A.) 355–390, (Butterworth-Heinemann, Stoneham, MA, 1992).
- Debets, F., Swart, K., Hoekstra, R.F. & Bos, C.J. Genetic maps of eight linkage groups of *Aspergillus niger* based on mitotic mapping. *Curr. Genet.* **23**, 47–53 (1993).
- Verdoes, J.C. et al. The complete karyotype of *Aspergillus niger*: the use of introduced electrophoretic mobility variation of chromosomes for gene assignment studies. *Mol. Gen. Genet.* **244**, 75–80 (1994).
- Adams, T.H., Wieser, J.K. & Yu, J.H. Asexual sporulation in *Aspergillus nidulans*. *Microbiol. Mol. Biol. Rev.* **62**, 35–54 (1998).
- Geiser, D.M., Timberlake, W.E. & Arnold, M.L. Loss of meiosis in *Aspergillus*. *Mol. Biol. Evol.* **13**, 809–817 (1996).
- Paoletti, M. et al. Evidence for sexuality in the opportunistic fungal pathogen *Aspergillus fumigatus*. *Curr. Biol.* **15**, 1242–1248 (2005).
- Bernard, M. & Latgé, J.P. *Aspergillus fumigatus* cell wall: composition and biosynthesis. *Med. Mycol.* **1** Suppl 9–17 (2001).
- Levin, D.E. Cell wall integrity signaling in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **69**, 262–291 (2005).
- Wösten, H.A.B. Hydrophobins: multipurpose proteins. *Annu. Rev. Microbiol.* **55**, 625–646 (2001).
- Askolin, S., Penttilä, M., Wösten, H.A.B. & Nakari-Setälä, T. The *Trichoderma reesei* hydrophobin genes hfb1 and hfb2 have diverse functions in fungal development. *FEMS Microbiol. Lett.* **253**, 281–288 (2005).
- Emri, T., Molnár, Z., Pusztahelyi, T., Varcza, Z. & Pócsi, I. The FluG-BrlA pathway contributes to the initiation of autolysis in submerged *Aspergillus nidulans* cultures. *Mycol. Res.* **109**, 757–763 (2005).



24. White, S., McIntyre, M., Berry, D.R. & McNeil, B. The autolysis of industrial filamentous fungi. *Crit. Rev. Biotechnol.* **22**, 1–14 (2002).
25. David, H., Akesson, M. & Nielsen, J. Reconstruction of the central metabolism of *Aspergillus niger*. *Eur. J. Biochem.* **270**, 4243–4253 (2003).
26. Förster, J., Famili, I., Fu, P., Palsson, B.O. & Nielsen, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**, 244–253 (2003).
27. Todd, R.B. & Andrianopoulos, A. Evolution of a fungal regulatory gene family: the Zn (II)Cys6 binuclear cluster DNA binding motif. *Fungal Genet. Biol.* **21**, 388–405 (1997).
28. Karaffa, L. & Kubicek, C.P. *Aspergillus niger* citric acid accumulation: do we understand this well working black box. *Appl. Microbiol. Biotechnol.* **61**, 189–196 (2003).
29. Pedersen, H., Hjort, C. & Nielsen, J. Cloning and characterization of *oah*, the gene encoding oxaloacetate hydrolase in *Aspergillus niger*. *Mol. Gen. Genet.* **263**, 281–286 (2000).
30. Poore, R.E., Hurst, C.H., Assimos, D.G. & Holmes, R.P. Pathways of hepatic oxalate synthesis and their regulation. *Am. J. Physiol.* **272**, C289–C294 (1997).
31. Schröder, M. & Kaufman, R.J. ER stress and the unfolded protein response. *Mutat. Res.* **569**, 29–63 (2005).
32. Steel, G.J., Fullerton, D.M., Tyson, J.R. & Stirling, C.J. Coordinated activation of Hsp70 chaperones. *Science* **303**, 98–101 (2004).
33. Travers, K.J. *et al.* Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell* **101**, 249–258 (2000).
34. van Huizen, R., Martindale, J.L., Gorospe, M. & Holbrook, N.J. P58^{IPK}, a novel endoplasmic reticulum stress-inducible protein and potential negative regulator of EIF2 α signalling. *J. Biol. Chem.* **278**, 15558–15564 (2003).
35. Römisch, K. Endoplasmic reticulum associated degradation. *Annu. Rev. Cell Dev. Biol.* **21**, 435–456 (2005).
36. Verma, R. *et al.* Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. *Mol. Biol. Cell* **11**, 3425–3439 (2000).
37. de Vries, R.P., Benen, J.A.E., de Graaff, L.H. & Visser, J. Plant cell wall degrading enzymes produced by *Aspergillus*. in *The Mycota X Industrial Applications* (ed. Osiewacz, H.D.) 263–279, (Springer, Berlin Heidelberg, 2002).
38. Adams, D.J. Fungal cell wall chitinases and glucanases. *Microbiology* **150**, 2029–2035 (2004).
39. Keller, N.P., Turner, G. & Bennett, J.W. Fungal secondary metabolism – from biochemistry to genomics. *Nat. Rev. Microbiol.* **3**, 937–947 (2005).
40. Proctor, R.H., Brown, D.W., Plattner, R.D. & Desjardins, A.E. Co-expression of 15 contiguous genes delineates a fumonisin biosynthetic gene cluster in *Gibberella moniliformis*. *Fungal Genet. Biol.* **38**, 237–249 (2003).
41. O'Callaghan, Caddick, .X. & Dobson, D.W. A polyketide synthase gene required for ochratoxin A biosynthesis in *Aspergillus ochraceus*. *Microbiology* **149**, 3485–3491 (2003).
42. Moss, M.O. Recent studies of mycotoxins. *J. Applied Microbiol. Symp. Suppl.* **84**, 62S–76S (1998).
43. Reiber, K. *et al.* The expression of selected non-ribosomal peptide synthetases in *Aspergillus fumigatus* is controlled by the availability of free iron. *FEMS Microbiol. Lett.* **248**, 83–91 (2005).
44. Salamov, A.A. & Solovyev, V.V. *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.* **10**, 516–522 (2000).
45. Frishman, D. *et al.* Functional and structural genomics using PEDANT. *Bioinformatics* **17**, 44–57 (2001).
46. Thompson, J.D., Plewniak, F., Thierry, J. & Poch, O. DbClustal: rapid and reliable global multiple alignments of protein sequences detected by database searches. *Nucleic Acids Res.* **28**, 2919–2926 (2000).
47. Schmidt, H.A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504 (2002).
48. Müller, T. & Vingron, M. Modeling amino acid replacement. *J. Comput. Biol.* **7**, 761–776 (2000).
49. Kumar, S., Tamura, K. & Nei, M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. *Brief. Bioinform.* **5**, 150–163 (2004).
50. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).