

SUPPLEMENTARY DATA

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 Breststraat¹, 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, A. Martin 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.I. van de Vondervoort²¹, Holger Wedler²⁴, Han A.B. Wösten⁷, An-Ping Zeng²⁵, Albert J.J. van Ooyen¹, Jaap Visser²⁹ and Hein Stam¹

¹ DSM Food Specialties, PO Box 1, 2600 MA Delft, The Netherlands

² Kluyver Centre for Genomics of Industrial Fermentation, Department for Biotechnology, Delft University of Technology, Julianalaan 67, 2628 BC Delft, The Netherlands

³ School of Biology, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

⁴ Center for Microbial Biotechnology, Technical University of Denmark, Building 223, Søtofts Plads, DK-2800 Kgs. Lyngby, Denmark

⁵ Section Fungal Genomics, Laboratory of Microbiology, Wageningen University, Dreijenlaan 2, 6703 HA Wageningen, The Netherlands.

⁶ Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, UK

⁷ Microbiology, Science Faculty, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

⁸ Biomax Informatics AG, Lochhamer Str. 9, 82152 Martinsried, Germany

⁹ CLC bio, Gustav Wieds Vej 10, 8000 Aarhus C, Denmark

¹⁰ DSM Anti Infection, PO Box 425, 2600 KA Delft, The Netherlands

¹¹ The University of Liverpool, School of Biological Sciences, Biosciences Bld., Crownstreet, Liverpool L69 7ZB, UK

¹² Ghent University and VIB, Dept. Molecular Biomedical Research, Unit of Fundamental and Applied Molecular Biology, Technologiepark 927, 9052 Gent, Belgium

¹³ Centre for the Analysis of Biological Complexity, Faculty of Life Sciences, The University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK

¹⁴ Architecture et Fonction des Macromolécules Biologiques, UMR6098, CNRS Universités Aix-Marseille I & II, Case 932, 163 Avenue de Luminy, 13288 Marseille (France)

¹⁵ Laboratory of Genetics, Wageningen University, Arboretumlaan 4, 6703 BD Wageningen, The Netherlands

¹⁶ Centre for Carbohydrate Bioprocessing (CCB) TNO-RuG, P.O. Box 14, 9750 AA Haren, The Netherlands

¹⁷ Department of Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute and Materials Science Center Plus, University of Groningen, P.O. Box 14, 9750 AA Haren, The Netherlands

¹⁸ Unité Postulante Biologie et Pathogénicité Fongiques, INRA USC 2019, Institut Pasteur, 75724 Paris Cedex 15, France

¹⁹ Swammerdam Institute for Life Sciences, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

²⁰ Laboratoire de Microbiologie, UMR 77, Pathologie Végétale, Université d'Angers, 49045 Angers Cedex, France

²¹ Institute of Biology Leiden, Leiden University, Molecular Microbiology, Wassenaarseweg 64 2333 AL Leiden, The Netherlands

²² 260 CMBI, Radboud University Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands

²³ Institute of Chemical Engineering, Research Area Gene Technology and Applied Biochemistry, Technical University Vienna, 1060 Vienna, Austria

²⁴ Qiagen Genomics Services, Qiagen GmbH., 40724 Hilden, Germany

²⁵ Helmholtz Center for Infection Research (former GBF- German Research Centre for Biotechnology) Inhoffenstrasse 7, D-38124 Braunschweig, Germany

²⁶ Department of Microbiology, Faculty of Sciences, University of Szeged, P.O.Box 533, H-6701 Szeged, Hungary

²⁷ Center for Biological Sequence Analysis, BioCentrum-DTU, Building 301, The Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

²⁸ CBS Fungal Biodiversity Centre, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

²⁹ Fungal Genetics and Technology Consultancy, PO Box 396, 6700 AJ Wageningen, The Netherlands

Correspondence should be addressed to Hein Stam

Hein.Stam@dsm.com

Genome sequencing and assembly

The *A. niger* genome was sequenced using a BAC walking approach as explained in the Methods. On average, a BAC end-sequence was found every 2.04 kb in the genome, indicating a genome size of approximately 36 Mb. Pulsed-field gel electrophoresis of *A. niger* strain N400 (CBS 120.49) chromosomes suggested a genome size of about approx. 37 Mb¹. Mapping the end-sequences onto the final genomic sequence showed a very even distribution of those end-sequences with no local clustering, underlining the good random cloning of large genomic sub-fragments into this BAC library. The only exceptions were clones and sequences relating to the rDNA region of the genome. There were no further large repetitive regions noticed. Smaller repeat regions have been resolved for each individual BAC. Further, no repeats within BAC/BAC overlapping regions, potentially confounding a correct BAC-to-BAC assembly, were found. In addition to the BAC-to-BAC assembly based on overlapping regions, all BAC end-sequences with their forward/reverse constraints per clone as well as sizing information for individual BAC clones were used to layer a BAC map on top of the resulting assemblies. The consistency of the assembly was checked on the back of that BAC map for each BAC/BAC overlap and assembly.

In total, 505 BACs have been sequenced with an average insert size of 76.8 kb and an

average 7.5-fold coverage. The assembled BACs contained on average 2.14 contigs per BAC corresponding with on average 1.14 sequence gaps per BAC. Six BACs have been identified as chimeric clones using BAC end-sequence information.

The genome was assembled and logically joined using clones physically bridging gaps to form 19 supercontigs with a unique total size of 33.9 Mb. Sequenced overlaps between individual BACs was 4.4 Mb (approximately 11.4% of the total sequence generated). The estimated 5% of the genome not yet sequenced includes telomeric regions, additional rDNA repeats, and small gaps. These results indicated that using end-sequencing as a way to map the BAC clones allowed for high accuracy and eventual direct alignment onto the assembled genomic contigs as well as sequence comparisons between all sequences obtained during the course of the project.

Establishing the orientation of supercontigs on chromosomes

Four different methods were used to establish the orientation of supercontigs on the chromosomes. First, we searched for telomeric repeats. In *A. nidulans* telomeric repeats of the hexamer TTAGGG were found², while for *A. oryzae* TTAGGGTCAACA repeats were found³. In our sequences we found only one telomeric repeat, consisting of a sixteen fold repeat of TTAGGGTT, where in two cases a T was missing. This telomeric repeat was located in BAC end EQ03278.p1, positioning it on the end of supercontig An04 on chromosome VI. Apparently this sequencing method is unfavorable for detecting telomeres.

A second method was to determine which supercontigs represent the left or the right arms. For three chromosomes we could make a direct link using the cloned genetic markers *niaD*, *nicB*, *nirA* and *pyrA*. Thus, the orientation of the corresponding supercontigs representing linkage groups III, VII and VIII could be concluded (**Table 2**).

A third method uses the high degree of synteny between *Aspergillus* species near the centromeres. For *A. nidulans* a high quality genetic map is available, which has been linked to the physical map⁴. Although synteny with *A. niger* was very weak near the telomeric regions of the *A. nidulans* scaffolds, synteny was remarkably high near the centromeres. On one side of 16 *A. niger* supercontigs there is a high degree of synteny with the *A. nidulans* scaffolds closest to the centromeres, representing eight pairs of supercontigs allocated to each linkage group, leaving three small supercontigs with unknown orientation.

Finally, microarray expression data underpin the proposed orientation of the supercontigs (not shown). The telomeric positioning effect, a reduced expression of genes located near the telomeres is clearly applicable for *A. niger* in the known telomeric ends of the supercontigs of chromosomes I, II, III, VI and VII. Looking at expression on either side of the three other supercontig pairs and the three unpositioned supercontigs, the telomeric ends could be identified and the proposed reciprocal orientation of all supercontigs is presented in **Table 2**.

***A. niger* genes relating to early sexual development**

A characteristic *MAT-1* family alpha-domain mating-type gene is closely linked to genes found flanking the *MAT* locus in other euascomycete fungi, although some syntenic rearrangement is observed compared to other aspergilli⁵. No *MAT-2* family HMG mating-type gene was detected. Hence, the *A. niger* isolate used for genome sequencing is of *MAT-1* genotype. Furthermore, a pheromone-precursor gene containing two internal repeats for an α -factor-like pheromone is present, together with a series of genes for enzymatic processing and maturation of **a**-factor and α -factor-like pheromones, and a gene for pheromone efflux. In contrast, the presence of an **a**-factor like pheromone-precursor gene could not be confirmed, although various candidate sequences were identified. Complementary α -factor and **a**-factor pheromones are required for sex to occur in heterothallic species, and the absence of an **a**-factor pheromone could provide an explanation for asexuality in *A. niger*. However, **a**-factor precursor genes are highly divergent and short in length⁶ and thus very difficult to detect. Indeed, it was not possible to find an **a**-factor like precursor in the known sexual species *A. nidulans*⁵.

In sexual species, pheromone binding to a cognate pheromone receptor leads to activation of a MAP-kinase signaling pathway⁷. The *A. niger* genome contains genes encoding characteristic G-protein coupled seven transmembrane domain receptors for both **a**- and α -factor-like pheromones. In addition, genes encoding alpha, beta and gamma G-protein subunits were found. The gamma-subunit gene was found as a partial 136 bp sequence at the direct start of a contig. Finally, the full complement for a MAP-kinase pathway, including elements linked specifically to sexual reproduction and a linked homeodomain transcriptional activator were detected (**Supplementary Table 5**).

A further 17 genes implicated in development of ascospores in other ascomycete species^{8,9} were detected in *A. niger*. These include genes required for the initiation of sexual morphogenesis, for Psi hormone production, for ascus development, for synthesis of elements of the COP9 signalosome, for general metabolism required during sexual development and genes for a series of transcription factors (**Supplementary Table 5**). All appear to encode functional proteins. Only the *pro1* gene equivalent which is required for development of protoascospores in *Sordaria*¹⁰ contained an apparent stop codon in the centre of the reading frame, and therefore may not be functional. In addition, only a partial *medA* gene with a truncated C-terminus is found at the direct end of a contig. The genome was not at this stage screened for genes involved with meiosis and ascospore formation.

The genes involved in heterokaryon incompatibility were never characterized at the molecular level in any *Aspergillus* species. A series of genes similar to genes involved in vegetative incompatibility in other organisms¹¹⁻¹³ is present in the *A. niger* genome, consistent with observations of heterokaryon incompatibility among natural isolates of *A. niger*¹⁴ (**Supplementary Table 6**). Manipulation of *het*-genes offers the possibility of utilizing the parasexual cycle, which has been previously used for genetic mapping¹⁵, for strain

improvement in a biotechnological setting.

Transport

Transport systems play essential roles in cellular metabolism, such as nutrient uptake, excretion of toxic compounds and secondary metabolites, maintenance of ion homeostasis but also in sensory processes¹⁶. Insight in the distribution of transport protein classes is vital to allow understanding of the metabolic capability of sequenced organisms¹⁷. Transport systems differ in their membrane topology and subunit composition, energy coupling mechanisms and substrate specificities¹⁶. Mostly, ATP and the electrochemical transmembrane gradient of sodium ions or protons are used to drive the transport processes. For *A. niger*, a complete set of membrane transport systems was identified with predicted functions, and classified into protein families based on the transporter classification system (<http://www.tcdb.org/>). This system represents a systematic approach to classify transport systems according to their mode of transport, energy coupling mechanism, molecular phylogeny, and makes predictions on substrate specificity. Herein, the transport mode and energy coupling mechanism serve as the primary basis for classification because of their relatively stable characteristics, whereas substrate specificity prediction is uncertain. There are three major classes of solute transporters found in *A. niger* within the transporter classification system: channels, primary (active) transporters, and secondary transporters. Transporters of unknown mechanism or function are included as a distinct class. Transporters involved in protein export and import were excluded from the analysis to allow focus on nutrient and ion transport only. Each transporter class is further classified into individual families and subfamilies according to their function, phylogeny, and/or substrate specificity.

A total of 865 transport proteins were predicted by our analysis of the *A. niger* genome. These were classified into 56 families, including 4 families of primary transporters, 39 families of secondary transporters, 10 channel protein families, and 3 unclassified families (**Supplementary Table 12**). With respect to genome size, a relatively high density of transporter genes is apparent in *A. niger*, a characteristic shared with *A. oryzae* (<http://www.tcdb.org/>). In particular, the major facilitator superfamily (MFS) with multiple transporter gene paralogs appears exceptionally large in both *A. niger* (461) and *A. oryzae* (507) with lower numbers in *A. fumigatus* (275) and *A. nidulans* (358) (**Supplementary Table 13**). MFS proteins have been shown to facilitate the transport of a diverse range of molecules including the following: di- and monosaccharides; polyols; quinate; inorganic phosphate; siderophores; drugs such as antifungals; mono- and dicarboxylates; and various other organic acids. MFS proteins catalyze uniport, solute:cation (H⁺ or Na⁺) symport and/or solute:H⁺ or solute:solute antiport. In addition, a few MFS proteins have been shown to function as glucose sensors. To date all characterized fungal sugar transporters belong to the MFS. The relatively high abundance of this transporter family in *A. niger* is consistent with its nutritional

versatility. Another large group of secondary transporters belong to the amino acid polyamine-organocation (APC) family. These transporters allow the cell factory to efficiently internalize the amino acids resulting from extracellular proteolytic degradation (**Supplementary Table 18**). Also, the ATP binding cassette (ABC) superfamily of transporters is relatively abundant in *A. niger* which includes members involved in multidrug resistance, long chain fatty acids and (pheromone) peptide transport. Several of the MDR-like ABC transporter genes are found in NRPS and PKS encoding gene clusters and are likely to be involved in the secretion of secondary metabolites.

Central metabolism

Despite a highly active glycolysis the genes encoding the enzymatic steps from glucose to pyruvate through the Embden-Meyerhof-Parnas and hexose monophosphate pathways did not reveal any exceptional features except for the presence of five hexokinases and four aldolases. Two of the putative hexokinase-related proteins may not function as such. A homolog (XprF) in *A. nidulans* regulates extracellular protease production upon carbon starvation¹⁸. Single genes are present for 6-phosphofructokinase and fructose-1,6-bisphosphatase (**Supplementary Figure 3**). Furthermore, single genes for 6-phosphofructokinase-2 (An07g02100) and fructose-2,6-bisphosphatase (An15g00200) involved in synthesis and degradation of the regulatory metabolite F2,6P were found. These enzymes bear consensus sequences for regulation by protein phosphorylation, similar to the situation in other organisms. Consequently, a full complement of the G-protein/cAMP signaling pathway was identified. Only the distribution of G-protein-coupled receptors (GPCRs) among the various structural and functional classes and the occurrence of an additional heterotrimeric G-protein alpha subunit and a second regulatory subunit of the cAMP-dependent protein kinase¹⁹⁻²¹ indicate differences with other fungi (**Supplementary Figure 4**). The last two enzymes of the Entner-Doudoroff pathway that distinguish it from the hexose monophosphate pathway were lacking, suggesting that this route is absent in *A. niger*. However, this finding should be considered with some caution as enzyme data have been reported in the past indicative for the presence of both gluconate dehydratase and 2-keto-3-deoxygluconate aldolase activities^{22,23}. Moreover, two genes with high similarity to 2-deoxy-D-gluconate 3-dehydrogenase (An04g02770 and An12g02700) were found.

A. niger appears to be well equipped for alcoholic fermentation with three genes encoding pyruvate decarboxylase and a large family of alcohol dehydrogenases. However, the functions of these enzymes in alcoholic fermentation or alcohol utilization are not clear and they may have various substrate specificities. A similar but smaller diversity of alcohol dehydrogenases has been observed in *N. crassa*²⁴.

A. niger comprises extensive metabolic pathways for protection against osmotic stress. The genome contains three genes encoding trehalose 6-phosphate synthase subunits, one for

trehalose 6-phosphate phosphatase and two for regulatory subunits. Interestingly, one trehalose 6-phosphate synthase gene (*tpsB*)²⁵ and one regulatory subunit gene are located adjacently, as in *A. fumigatus* and *A. oryzae*. The absence of these two genes in *A. nidulans*^{5,26,27} may explain the phenotypic differences that have been observed between *A. nidulans* and *A. niger* mutants lacking one of the T6P synthases^{25,28}. The high-osmolarity compatible solute glycerol can be synthesized through two parallel pathways²⁹ (**Supplementary Figure 3**). Genes encoding components of both pathways were identified except for a postulated DHAP phosphatase. Hence, we speculate that the two glycerol 3-phosphate phosphatases of *A. niger* may also serve the function of DHAP phosphatases as previously suggested for *A. nidulans*³⁰. Genes for uptake of glycerol, a glycerol kinase and a mitochondrial FAD-dependent glycerol 3-phosphate dehydrogenase were identified except the small and membrane bound subunits of the latter enzyme³¹. Genes encoding efflux pumps related to *S. cerevisiae* Fps1 and to aquaporins³² were also identified. These may serve an important role in regulating intracellular glycerol levels (**Supplementary Figure 3**).

The *A. niger* protein secretion machinery

A. niger encodes three soluble luminal protein disulphide isomerases and one putative membrane-bound PDI-family protein (EpsA). Although the same number as in yeast, only PdiA (Pdi1p) and EpsA (Eps1p) appear to be close homologs of yeast proteins. TigA³³ and PrpA³⁴ are not direct homologs of yeast Mpd1p or Mpd2p. PdiA, TigA and EpsA have two thioredoxin domains each whereas PrpA has only one. The domains in EpsA are not the typical CGHC as in the others (being CPHC and CHHC) so functionality is not assured. Indeed, yeast Eps1p was shown not to have detectable isomerase activity³⁵ and Pdi1p has more activity than the other foldases in yeast³⁶.

Interestingly, the *A. niger* ER luminal protein EroA has a predicted C-terminal ER-retention signal (HEEL) as seen in *A. nidulans* (RDEL), *A. fumigatus* (HDEL) and *N. crassa* (LDEL), whereas in several yeast species, plants and mammals no such retention signal is present. Retention in the ER is due to a so far unspecified interaction with a membrane protein³⁷. In contrast, the EvrA ortholog of yeast Erv2p, has no predicted ER retention sequence.

An additional peptidyl prolyl isomerase is apparently present in the ER lumen of *A. niger*, in addition to CypB³⁸. Moreover, a calnexin and putative luminal HSP70s are present including BipA and LhsA. Kar2p (equivalent to *A. niger* BipA) and Lhs1p function together in yeast alongside DnaJ proteins and the nucleotide exchange factor Sil1p³⁹. Surprisingly, no homolog of Sil1p can be found in *A. niger* and other aspergilli. Although the degree of similarity between Sil proteins from various sources is low except for one region (Colin Stirling, personal communication), there are readily detectable putative homologs in *N. crassa*.

The Unfolded Protein Response (UPR) signaling pathway seems to be yeast-like in its gene complement except for the presence of a putative ortholog of mammalian p58 of the PERK/eIF2alpha/ATF4 pathway which is absent in yeast⁴⁰⁻⁴². P58 is involved in translational

regulation, attenuating the UPR during ER stress in mammalian cells⁴³. No homologs of PERK or ATF4/6 are apparent in *A. niger*, hence the role of the putative p58 in *A. niger* remains intriguing.

ER-associated protein degradation (ERAD) which degrades misfolded or unassembled proteins at the proteasome following dislocation from the ER, is largely yeast-like with some differences. Putative homologs to several membrane proteins which seem to be essential players in ERAD in yeast, have a reduced similarity to predicted *A. niger* orthologs (similarity to Der1p, Hrd1p, Doa10p, Hrd3p). No homologs are found to several proteins known to be involved in ERAD in *S. cerevisiae*: (i) Cue1p, a membrane protein that recruits the ubiquitin-conjugating enzyme Ubc7p to the ER.; (ii) Rad23p, associated with the Cdc48p-Npl4p-Ufd1p complex involved in binding ubiquitinated proteins prior to their degradation; (iii) Ubx2p, of the family of UBX (ubiquitin-regulatory x)-domain containing proteins that bind ubiquitinated proteins; (iv) Yos9p, a receptor that recognizes misfolded N-glycosylated proteins and participates in their targeting to ERAD⁴⁴. Homologs to all the subunits of the 26S proteasome are found whereas homologs of the regulatory proteins Rpn13 and Rpn14, subunits of the 19S regulatory particle of the 26S proteasome lid⁴⁵, appear to be either absent or weakly similar in *A. niger* and other filamentous fungi.

Sugar nucleotides act as activated monosaccharide donor substrates for the synthesis of oligosaccharide structures in the ER and Golgi. *A. niger* contains several homologs of genes involved in the synthesis of UDP-Glc, UDP-GlcNAc, GDP-Man and UDP-Gal. Furthermore, two homologs of UDP-Galp mutase⁴⁶ are present, enabling the organism to synthesize UDP-Galf. Genes involved in the transport of sugar nucleotides into the lumen of the ER or Golgi apparatus, were also identified. Almost all genes involved in the synthesis of the lipid-linked oligosaccharide precursor Glc₃Man₉GlcNAc₂ and in its transfer to the growing polypeptide, could be identified in the *A. niger* genome. One exception is *ALG14*, although homologs of this gene were found in *A. fumigatus* and *A. nidulans*. Interestingly, the dolichol-P-mannose synthase gene is more mammalian- than yeast-like since it does not contain a C-terminal hydrophobic transmembrane region. One glucosidase I and several genes with homology to the catalytic alpha subunit of glucosidase II were found. However, some of the latter might also encode glucosidases that may be secreted into the medium, depending on the growth conditions. One of those has a very high similarity to *S. cerevisiae* Rot2p and to the GlcIIa from *T. reesei*⁴⁷. One homolog was found to the glucosidase II beta subunit and the ER soluble UDP-Glc:glycoprotein glucosyltransferase. Hence, *A. niger*, potentially possesses a well-developed glycosylation-dependent quality control system. Several homologs of the yeast ER mannosidase Mns1p were identified, which trims the Man₉GlcNAc₂ structure to Man₈GlcNAc₂. One of those shows high similarity to the yeast Htm1p/EDEM homolog, involved in directing unfolded proteins to the ERAD pathway⁴⁸.

Although hyperglycosylation is rare in *A. niger*, in some cases the N-glycans on secreted proteins can contain up to 18 or more hexose residues^{49,50} due to the activity of

glycosyltransferases within the Golgi apparatus. Accordingly, many genes encoding Golgi-typical putative type II mannosyltransferases are present. The presence of two homologs of yeast Mnn4p confirms the potential of the organism to synthesize fungal-specific phosphomannose residues. Galactose residues in the furan form are often found on N- as well as O-glycans of *A. niger*⁵¹⁻⁵⁴. However, a galactofuranosyl transferase could not be identified.

O-glycans in *A. niger* are of the O-mannose type and small with, typically, 1 to 3 residues. They may contain alpha-1,2 and alpha-1,6-mannoses as well as glucose and galactofuranose residues and can be branched^{50,55}. The *A. niger* genome contains three protein O-mannosyltransferase (*PMT*) family homologs which are highly similar multi-membrane-spanning proteins that transfer the first mannose to the Ser/Thr residue of the polypeptide. In addition, several homologs of mannosyltransferases are found (Ktr1p, Mnn2p, Mnt2p), but as yet it is unclear whether they act on N- or O-glycans or on both. Expression data (**Supplementary Table 7**) reveal that in all cases one to several of the redundant genes for glycosylation are transcribed in standard fed-batch conditions.

Of the secretion-associated small GTPases the Sec4p homolog in *A. niger* was shown not to be essential⁵⁶. A total of ten secretion-related GTPase-encoding genes is present in the *A. niger* genome with varying similarity to mammalian or to yeast counterparts. No homologs to YPT10/11 could be found. Vesicles moving from the ER to Golgi are CopII-coated. Whereas *S. cerevisiae* contains eight coat protein genes, only seven homologs are present in *A. niger*. A homolog of Sec28p seems to be present in neither *A. niger* nor *A. nidulans*, *A. fumigatus* or *N. crassa*. Vesicles involved in retrograde transport from the Golgi to the ER are COPI-coated and all COPI coat proteins have clear homologs in *A. niger*. Several genes are present encoding v-Snare and t-Snare proteins which interact in vesicle fusion at the receiving secretion compartment. Two highly similar homologs with t-Snare activity and three with v-Snare activity were identified, together with nine homologs with moderate similarity to *S. cerevisiae* v- or t-Snares. Intriguingly, *A. niger* appears to contain nine v- and t-Snare genes less than *S. cerevisiae*.

For vesicles moving from the ER to Golgi a conserved oligomeric tethering complex (Cog1 through Cog8) is involved in transporting and docking of the vesicles. In *A. niger* clear homologs of Cog3 and Cog4 are found and homologs with weak similarity are present for Cog5 and Cog6. Similar results were found for *A. nidulans*, *A. fumigatus* and *N. crassa*. The transport protein particle (TRAPP) complex mediates vesicle docking and fusion at the cis-Golgi. Eight TRAPP-protein genes with moderate similarity were found whereas no homologs could be identified for Trs20p and Kre11p. The exocyst complex is required for the docking of secretory vesicles to the plasma membrane. Homologs of all eight proteins of the *S. cerevisiae* exocyst are found in *A. niger*, together with homologs of the three small GTPases interacting with this complex (Rho1p, Cdc42p and Sec4p).

Polysaccharide degrading enzymes

Fungal cellulases, hemicellulases, pectinases, amylases and inulinases are used in a wide range of industrial applications, for example in increasing filterability in brewing⁵⁷, in the biobleaching of pulps^{58,59}, increasing bread volume and dough quality in baking^{60,61}, and the conversion efficiency in animal feeds⁶², in clarifying juices⁶³, and are used to reduce the amount of oligosaccharides in soybean milk and sugar beet syrup^{64,65}. Depending on the application, specific (sets of) enzymes are required. *A. niger* is one of the most frequently used organisms to produce these enzymes and therefore well studied, particularly with respect to plant polysaccharide degradation⁶⁶⁻⁶⁸. The availability of genome sequences for *A. niger*, *A. nidulans*⁵, *A. oryzae*²⁶ and *A. fumigatus*²⁷ allows a comparison of these aspergilli with respect to their hydrolytic potential to degrade different kinds of biopolymers which they utilise in their natural environment as substrates. The carbohydrate-active enzymes content of *A. niger* has been compared against a set of more than 10 fungal genomes recently annotated in the Carbohydrate-Active enZYmes (CAZy) database (<http://afmb.cnrs-mrs.fr/CAZY/>)⁶⁹.

Here we analysed the *Aspergillus* genomes for the presence of ORFs that could be assigned to the different CAZy families (**Table 3**). The number of identified ORFs was found to be similar in all four species. However, the *A. niger* genome contains a significantly higher number of ORFs than the other aspergilli for CAZy family GH28 and also contains the highest number of ORFs belonging to family GH13. Enzymes within these families are involved in the degradation of pectin (GH28) or in the case of family GH13 in starch degradation, alpha 1,3-glucan synthesis and glycogen debranching.

Glycosyl hydrolase family 28

Pectins are complex heteropolysaccharides which occur in the middle lamella and primary cell wall of plants. These molecules have long been considered to form an extended backbone composed of smooth regions of homogalacturonan (partially methyl-esterified and O-acetylated) and of "hairy" regions consisting of rhamnogalacturonan I (RGI)⁷⁰.

Homogalacturonan has recently been proposed to be a side chain of RGI⁷¹. The latter structure is decorated by neutral sugars, usually (polymers of) arabinosyl and/or galactosyl residues⁷⁰. Other pectin structures described are xylogalacturonan and rhamnogalacturonan II which appear as decorations on homogalacturonan^{70,72}. To degrade this complex substrate, fungi require the combined action of a variety of enzymes. Two enzymatic mechanisms are known to cleave the main chains in the different pectin structures. One involves hydrolysis catalysed by hydrolases with different specificities, the other involves transelimination catalysed by various lyases. Many of the enzymes concerned and their corresponding genes

have been characterised in the past⁶⁶. The hydrolases are comprised of endo- and exo-polygalacturonases, endo- and exo-rhamnogalacturonan hydrolases, xylogalacturonan hydrolases and rhamnogalacturonan rhamnhydrolases/ α -rhamnosidases. They all belong to the GH28 family whereas the lyases belong to different families viz. PL1 (pectin lyases), PL1,3,9 (pectate lyases) and PL 4,11 (rhamnogalacturonan lyases). Also relevant are pectin methyl esterases (CE8) and rhamnogalacturonan acetyl esterases, (CE12) which are known to influence the action of the main chain cleaving enzymes^{73,74}. Comparative analysis with other fungi shows that the GH28 family is particularly enriched in the aspergilli (data not shown). ORFs assigned to GH28 from the four *Aspergillus* genomes were therefore analysed in more detail to assign enzymatic functions. This analysis was based on similarity to already characterised genes encoding GH28 enzymes of known function, and on the presence of specific residues in the amino acid sequence which in the case of polygalacturonase II (PGII) have been shown to be involved in substrate binding or catalysis^{75,76}. The presence or absence of specific signatures was determined for all GH28 members of the four aspergilli. In particular the presence of a specific Arg and His residue (in *A. niger* PGII at position 256 and 223, respectively) is diagnostic for activity on homogalacturonan allowing polygalacturonases and rhamnogalacturonases to be distinguished. Endo-rhamnogalacturonan hydrolases characteristically contain a conserved catalytic Glu residue instead of an Asp (located at the comparable Asp 202 position of the conserved Asp201 Asp 202 residues of PGII)^{77,78}. In a recent study the exo-acting members of the GH28 family were characterized, resulting in the discovery of three exo-polygalacturonases and three putative exo-rhamnogalacturonases⁷⁹. The enzyme activity had been discovered before⁸⁰ but genes for the latter enzymes not. This analysis results in GH28 ORFs being assigned to the various enzyme functions as indicated in **Table 3**.

A. niger contains a significantly higher number of endo-polygalacturonases (7) than the other aspergilli (3-4). The same trend is observed for the endo-rhamnogalacturonases (6) with the exception of *A. oryzae*. In contrast, the number of pectate lyases in the *A. niger* genome (only one PL1 pectate lyase and no PL3 or PL9 pectate lyases) is much lower than for the other *Aspergillus* genomes (**Table 3**). The pectin lyase family is large in all cases. A slight difference is also found in the number of GH88 protei family. These enzymes are believed to hydrolyse 4,5-unsaturated galacturonic acid oligosaccharides, which are formed by the action of pectate lyases and pectin lyases The *A. niger* genome contains only one of these ORFs, while the other *Aspergillus* genomes contain two or three. Taken as whole, these observations suggest that *A. niger* uses a different enzyme repertoire to degrade pectin than the other described aspergilli. A possible reason for this difference is that *A. niger* acidifies its environment much more strongly than *A. nidulans*, *A. fumigatus* and *A. oryzae*. Pectate lyases have very low or no activity at all at acidic pH and would therefore not be efficient enzymes for an acidifying fungus, while polygalacturonases and pectin lyases are active at

(slightly) acidic pH. It is also likely that *A. niger* degrades rhamnogalacturonan backbones by hydrolysis. A further difference in pectin degradation between *A. niger* and the other three aspergilli relates to the degradation of the arabinan side chains. Unlike the other *Aspergillus* genomes the *A. niger* genome contains no exo-arabinanases (family GH93), indicating that *A. niger* relies on the combined action of endo-arabinanase and α -L-arabinofuranosidase for the hydrolysis of the arabinan side chains.

Glycosyl hydrolase family 13

The GH13 glycosyl hydrolase family is enriched in the aspergilli compared to other fungi. The annotation and analysis of the ORFs from family GH13 revealed the presence of a diverse group of enzymes homologous to α -amylases. Based on phylogenetic clustering, three separate groups of enzymes can be distinguished (**Supplementary Figure 6**). The first consists of four α -amylases which are secreted extracellularly, and play a role in the degradation of starch (**Table 3**).

A second group consists of ORFs that do not contain a signal sequence and therefore appear to be intracellular enzymes. This group of enzymes is evolutionary related to bacterial liquefying amylases rather than to the fungal extracellular amylases described thus far. Representative members are present in all *Aspergillus* genomes, and it is therefore very likely that these intracellular amylase-like enzymes have a specific but thus far unknown role in fungal metabolism.

A third group of ORFs are most likely anchored to the cell wall by a glycosyl-phosphatidylinositol (GPI)-anchor, because their amino acid sequences contain a C-terminal GPI-anchor attachment site⁸¹. It is known that GPI-anchored β -glucanases play a role in the maintenance of the cell wall containing β -glucan⁸². The clustering of An00g07038 and An00g07039 with genes encoding putative α -glucan synthases suggests an involvement in α -glucan metabolism. The conserved domains of the GPI-anchored α -glucanases differ slightly but clearly from the consensus sequence for the α -amylase family, with Asn, Gln or Asp replacing one or two highly conserved His residues involved in substrate binding. These histidines are conserved throughout the whole α -amylase family, except in a few cases where enzymes have different enzymatic activity, for example in the case of acarviosyltransferase⁸³. Similar divergence from the consensus sequence are also found in other annotated fungal amylases predicted to be GPI-anchored, suggesting the presence of a group of enzymes with specific biochemical activity and conserved function in the cell. The combination of cell wall location and aberrant conserved motifs might indicate that these enzymes process cell wall polymers such as nigeran, a glucan with alternating α -1,4- and α -1,3-glycosidic bonds which are not normally substrates for GH13 enzymes⁸⁴.

Other glycosyl hydrolase families

Other large glycosyl hydrolase families in the aspergilli are the GH 3 (17-23 members, predominantly beta-glucosidases), GH 31 (7-11 members, predominantly alpha-glucosidases), GH 43 (10-20 members) and GH 78 families (8 members, all alpha-rhamnosidases). A significant number of the *A. niger* ORFs present in the CAZy families do not contain a predicted signal sequence (**Supplementary Table 17**). Of those only one ORF seems to encode an endo-acting enzyme, a putative endo-galactanase (GH53); all others encode exo-acting enzymes and are most likely involved in the degradation of di- or oligosaccharides or glycosides imported by the fungus, or of oligosaccharides which are produced intracellularly. A substantial number of these enzymes are classified as beta-glucosidase (GH 3; 5 members) or alpha-rhamnosidase (GH 78; 4 members), but of the 7 alpha-glucosidases (GH 31) and 7 alpha-galactosidases (GH 27 and 36) there are only two and one enzymes, respectively, lack a signal sequence. While no evidence exists for oligosaccharide transport in *Aspergillus*, transport of the disaccharides maltose and lactose has been reported in the yeasts *S. cerevisiae*⁸⁵ and *Kluyveromyces lactis*⁸⁶, respectively. Transporters involved in sugar uptake are nearly all members of the Major Facilitator Superfamily and candidate disaccharide transporters have been identified in *A. niger* (**Supplementary Data; Supplementary Table 12**). The large number of different intracellular exo-acting glycosidases detected in the *Aspergillus* genomes reflects the ability of these fungi to metabolize a wide variety of glycosidic structures present as distinct metabolites or in the cell wall structures of plants.

SUPPLEMENTARY METHODS

Construction of the genomic BAC library

Aspergillus niger strain CBS 513.88 was used as the DNA donor. For the construction of the genomic BAC library of *A. niger*, the vector pBeloBAC 11 was used as described⁸⁷. *A. niger* cells from a 50 ml YPD (1% yeast extract, 2% peptone, 2% glucose) culture were washed twice with TSE buffer (25 mM Tris-HCl, 300 mM sucrose, 25 mM EDTA, pH 8) and resuspended in TSE buffer. Then, agarose plugs from these cells were prepared according to the Bio-Rad manual of the Chef DR II pulsed-field gel electrophoresis system (PFGE system) using 1.5% low melting point agarose. Pre-electrophoresis was carried out on a Bio-Rad PFGE system. Partial digestion of genomic DNA was carried out using *Sau3AI* for restriction⁸⁷. Gel electrophoresis was carried out on a Bio-Rad PFGE system under the following conditions: 6 V/cm, 90 s pulse, 13°C 18 h. Agarose digestion with gelase, ligation and transformation was carried out using the protocol mentioned⁸⁷. Subsequent electroporation of DH10B cells (Invitrogen) was again carried out according to the same protocol, and bacteria were plated onto 2YT plates supplemented with chloramphenicol as selecting agent. Clones obtained from that procedure were picked and used to inoculate 1.2

ml of 2YT supplemented with chloramphenicol. These bacterial cultures were used to prepare glycerol stocks in 96-well microtitre plate format as resource for all subsequent work.

Construction of shotgun libraries from BAC DNA

Large-scale preparations of BAC DNA were carried out using the Large-Construct kit from Qiagen (Qiagen GmbH, Hilden, Germany). After sonification and enzymatic repair of the ends, fragments of desired size (usually 1.2 - 1.5 kb) were isolated from a 1% preparative agarose gel using the MinElute Gel Extraction kit (Qiagen) and inserted into a *Sma* I-digested and alkaline phosphatase-treated pUC19 vector⁸⁸. Ligation was carried out with the Rapid Ligation kit (Roche) according to the manufacturer's protocol. The ligation mixture was then desalted using a QIAquick kit (Qiagen) according to the instructions of the supplier with the exception of the elution step performed with distilled H₂O. 1/10 volume of the eluted DNA was used for transformation of competent *Escherichia coli* DH10B cells using a Genepulser II device (Bio-Rad). 1 ml Luria Bertani (LB) medium was added and incubated for 1 h at 37°C. 1/200 and 1/20 volumes of the transformed cells were plated onto Petri dishes containing LB agar, ampicillin, X-Gal and isopropylthiogalactoside (IPTG)⁸⁸ and grown overnight at 37°C to determine the yield of recombinant clones. Usually the transformation frequency exceeded 10⁸ transformants per µg vector DNA and the white:blue ratio was approximately 10:1 or better.

DNA sequencing and DNA assembly

For subsequent DNA sequencing, plasmid DNA from white colonies was isolated from cultures grown in 1.2 ml 2YT containing ampicillin for 24 h at 37°C by shaking at 220 rpm. Plasmid purification of shotgun clones was carried out using the REAL Prep 96 kit (Qiagen). DNA sequencing reactions were set up using BigDye Terminator v 2.0 cycle sequencing chemistry (Applied Biosystems) and purified using DyeEx 96 (Qiagen). Sequencing data were generated using ABI Prism 3700 sequence analyzers. Base calling and quality checks were carried out using Phred⁸⁹. BAC assemblies and raw data were visualized and edited using the STADEN package (version 4.5; http://www.mrc-lmb.cam.ac.uk/pubseq/staden_home.html).

Gene identification and annotation

Analysis and annotation of the genomic sequences of *A. niger* was performed with a combined automatic and manual approach. Genes were predicted by a version of FGENESH⁹⁰ trained on known *A. niger* genes and genes of related organisms. In addition GeneMark⁹¹, GENSCAN⁹² and GeneWise⁹³ were used. FGENESH, GeneMark and GENSCAN were all three run on the entire genomic sequence to provide an initial set of predicted genes. Preference was given to FGENESH genes, for regions without any FGENESH prediction. GeneMark or GENSCAN models were extracted with preference for the GeneMark models. A

test set of 65 known *A. niger* proteins was used to evaluate the quality of automatic gene identification by FGENESH. Of the 65 proteins evaluated 62 proteins (94%) were positively identified and the gene model of 43 proteins (66%) was fully correct. For the annotation of the full genome all automatically predicted ORFs were manually curated on the basis of Blastp alignments and the predictions made by the other algorithms. This led to the modification of 5681 (40%) ORFs. In addition, the genomic sequence was also searched against the non-redundant protein database using Blastx⁹⁴. For all initially predicted genes a Blastp⁹⁴ analysis against a non-redundant protein database was performed. Based on the Blastp results for each gene GeneWise was run against the best blast matches. The gene models of the initially predicted genes were manually adjusted in case that the Blastp and GeneWise alignment indicated a suboptimal gene model.

For regions without any gene prediction with one of the three algorithms but with a significant Blastx match, genes were manually extracted by usage of the respective GeneWise alignment. Incomplete GeneWise protein alignments were extended to the first exon upstream to the nearest start codon, and the last exon downstream to the first stop codon.

Transfer RNAs were identified using the tRNAScan-SE program⁹⁵. Ribosomal RNAs were identified by Blastn against a database of all publicly available rRNA sequences.

Transcriptional analysis

Biomass samples from fermentations were directly frozen into liquid nitrogen and stored at 80 C.. Under liquid nitrogen grinded mycelium was treated with Trizol and chloroform. Total RNA was further isolated using the RNA easy kit (Qiagen). Concentration of total RNA was determined by spectrophotometry (A260). Quality and integrity of RNA were checked with the A260/A280 ratio and on the Agilent 2100 Bioanalyser. Probe synthesis and fragmentation were performed according to Affymetrix protocol. The probe synthesis was performed using the Bioarray High Yield RNA transcript labeling kit from Enzo. Hybridisation, washing, staining and scanning were done according to Affymetrix protocol (Affymetrix, inc. "GeneChip Expression Analysis Technical Manual, august 2002).

Micro Array Suite (MAS 5.0 Affymetrix) software was used for data extraction. *Spotfire Decision Site for functional genomics 7.1* was used for data analysis.

References

1. 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).
2. Bhattacharyya, A. & Blackburn, E.H. *Aspergillus nidulans* maintains short telomeres throughout development. *Nucleic Acids Res.* **25**, 1426-1431 (1997).
3. Kusumoto, K.I., Suzuki, S. & Kashiwagi, Y. Telomeric repeat sequence of *Aspergillus oryzae* consists of dodeca-nucleotides. *Appl. Microbiol. Biotechnol.* **61**, 247-251 (2003).

4. Clutterbuck, A.J. The validity of the *Aspergillus nidulans* linkage map. *Fungal Genet. Biol.* **21**, 267-277 (1997).
5. Galagan, J.E. *et al.* Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**, 1105-1115 (2005).
6. Coppin, E., de Renty, C. & Debuchy, R. The function of the coding sequences for the putative pheromone precursor in *Podospora anserina* is restricted to fertilization. *Eukaryot. Cell* **4**, 407-420 (2005).
7. Elion, E.A. Pheromone response, mating and cell biology. *Curr. Opin. Microbiol.* **3**, 573-581 (2000).
8. Dyer, P.S., Ingram, D.S. & Johnstone, K. The control of sexual morphogenesis in the Ascomycotina. *Biol. Rev.* **67**, 421-458 (1992).
9. Braus, G.H., Krappman, S. & Eckert, S.E. in *Molecular Biology of Fungal Development* (ed. Osiewacz, H.D.) 215-244 (Marcel Dekker Basel, 2002).
10. Masloff, S., Jacobsen, S., Pöggeler, S. & Kück, U. Functional analysis of the C6 zinc finger *pro1* involved in fungal sexual development. *Fungal Genet. Biol.* **36**, 107-116 (2002).
11. Glass, N.L. & Kaneko, I. Fatal attraction: nonself recognition and heterokaryon incompatibility in filamentous fungi. *Eukaryot. Cell* **2**, 1-8 (2003).
12. Dementhon, K., Saupe, S.J. & Clavé, C. Characterization of IDI-4, a bZIP transcription factor inducing autophagy and cell death in the fungus *Podospora anserina*. *Mol. Microbiol.* **53**, 1625-1640 (2004).
13. Pinan-Lucarre, B., Paoletti, M., Dementhon, K., Couлары-Salin, B. & Clavé, C. *Mol. Microbiol.* **47**, 321-333 (2003).
14. van Diepeningen, A.D., Debets, A.J.M. & Hoekstra, R.F. Heterokaryon incompatibility blocks virus transfer among natural isolates of black aspergilli. *Curr. Genet.* **32**, 209-217 (1997).
15. Debets, A.J.M., 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).
16. Saier, M.H. Jr. A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol Mol Biol Rev* **64**, 354-411 (2000).
17. Ren, Q., Paulsen, I.T. Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. *PLoS Comp Biol.* **1**, e27 (2005).
18. Katz, M.E., Masoumi, A., Burrows, S.R., Shirliff, C.G. & Cheetham, B.F. The *Aspergillus nidulans xprF* gene encodes a hexokinase-like protein involved in the regulation of extracellular proteases. *Genetics* **156**, 1559-1571 (2000).
19. Lafon, A., Han, K.H., Seo, J.A., Yu, J.H. & d'Enfert, C. G-protein and cAMP mediated signaling in aspergilli: a genomic perspective. *Fungal Genet. Biol.* **43**, 490-502 (2006).
20. Kulkarni, R.D., Thon, M.R., Pan, H. & Dean, R.A. Novel G-protein-coupled receptor-like proteins in the plant pathogenic fungus *Magnaporthe grisea*. *Genome Biol.* **6**, R24 (2005).
21. Dean, R.A. *et al.* The genome sequence of the rice blast fungus *Magnaporthe grisea*.

- Nature* **434**, 980-986 (2005).
22. Elzainy, T.A., Hassan, M.M. & Allam, A.M. New pathway for nonphosphorylated degradation of gluconate by *Aspergillus niger*. *J. Bacteriol.* **114**, 457-459 (1973).
 23. Allam, A.M., Hassan, M.M. & Elzainy, T.A. Formation and cleavage of 2-keto-3-deoxygluconate by 2-keto-3-deoxygluconate aldolase of *Aspergillus niger*. *J. Bacteriol.* **124**, 1128-1131 (1975).
 24. Borkovich, K.A. *et al.* Lessons from the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism. *Microbiol. Mol. Biol. Rev.* **68**, 1-108 (2004).
 25. Wolschek, M.F. & Kubicek, C.P. The filamentous fungus *Aspergillus niger* contains two differentially regulated trehalose-6-phosphate synthase-encoding genes, *tpsA* and *tpsB*. *J. Biol. Chem.* **272**, 2729-2735 (1997).
 26. Machida, M. *et al.* Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* **438**, 1157-1161 (2005).
 27. Nierman, W.C. *et al.* Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* **438**, 1151-1156 (2005).
 28. Fillinger, S. *et al.* Trehalose is required for the acquisition of tolerance to a variety of stresses in the filamentous fungus *Aspergillus nidulans*. *Microbiol.* **147**, 1851-1862 (2001).
 29. Yancey, P.H., Clark, M.E., Hand, S.C., Bowlus, R.D. & Someron, G.N. Living with water stress: evolution of osmolyte systems. *Science* **217**, 1214-1222 (1982).
 30. Hondmann, D.H.A. & Visser, J. in *Aspergillus: 50 years on* (eds Martinelli, S.D. & Kinghorn, J.R.) *Progr. Industr. Microbiol.* **29**, 61-139 (Elsevier Amsterdam 1994).
 31. Hondmann, D.H.A., Busink, R., Witteveen, C.F. & Visser, J. Glycerol catabolism in *Aspergillus nidulans*. *J. Gen. Microbiol.* **137**, 629-636 (1991).
 32. Hohmann, S. Osmotic adaptation in yeast- control of the yeast osmolyte system. *Int. Rev. Cytol.* **215**, 149-187 (2002).
 33. Jeenes, D.J., Pfaller, R. & Archer, D.B. Isolation and characterization of a novel stress-inducible PDI-family gene from *Aspergillus niger*. *Gene* **193**, 151-156 (1997).
 34. Wang, H. & Ward, M. Molecular characterization of a PDI-related gene *prpA* in *Aspergillus niger* var. *awamori*. *Curr. Genet.* **37**, 57-64 (2000).
 35. Wang, Q. & Chang, A. Eps1, a novel PDI-related protein involved in ER quality control in yeast. *EMBO J.* **18**, 5972-5982 (1999).
 36. Kimura, T. *et al.* Interactions among yeast protein disulfide isomerase proteins and endoplasmic reticulum chaperone proteins influence their activities. *J. Biol. Chem.* **280**, 31438-31441 (2005).
 37. Pagani, M., Pilati, S., Bertoli, G., Valsasina, B. & Sitia, R. The C-terminal domain of yeast Ero1p mediates membrane localization and is essential for function. *FEBS Lett.* **508**, 117-120 (2001).
 38. Derkx, P.M.F. & Madrid, S.M. The foldase CYPB is a component of the secretory pathway

- of *Aspergillus niger* and contains the endoplasmic reticulum retention signal HEEL. *Mol.Genet.Genom.* **266**,537-545 (2001).
39. Steel,G.J., Fullerton,D.M., Tyson,J.R. & Stirling,C.J. Coordinated activation of Hsp70 chaperones. *Science* **303**,98-101 (2004).
 40. Schröder,M. & Kaufman,R.J. ER stress and the unfolded protein response. *Mutation Res.***569**,29-63 (2005).
 41. Schröder,M. & Kaufman,R.J. The mammalian unfolded protein response. *Ann.Rev.Biochem.* **74**,739-789 (2005).
 42. 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).
 43. 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 EIF2alpha signalling. *J.Biol.Chem.***278**,15558-15564 (2003).
 44. Bhamidipati,A., Denic,V., Quan,E.R. & Weissman,J.S. Exploration of the topological requirements of ERAD identifies Yos9p as a lectin sensor of misfolded glycoproteins in the ER lumen. *Mol.Cell* **19**,741-751 (2005).
 45. 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).
 46. Bakker,H., Kleczka,B., Gerardy-Schahn,R. & Routier,F.H. Identification and partial characterization of two eukaryotic UDP-galactopyranose mutases. *Biol.Chem.* **386**,657-661 (2005).
 47. Geysens *et al.* Cloning and characterization of the glucosidase II alpha subunit gene of *Trichoderma reesei* : a frameshift mutation results in the aberrant glycosylation profile of the hypercellulolytic strain Rut-C30. *Appl.Environm. Microbiol.* **71**, 2910-2924 (2005).
 48. Jakob,C.A. *et al.* Htm1p, a mannosidase-like protein, is involved in glycoprotein degradation in yeast. *EMBO Reports* **2**,423-430 (2001).
 49. Warren,M.E. *et al.* Studies on the glycosylation of wild-type and mutant forms of *Aspergillus niger* pectin methylesterase. *Carbohydr. Res.* **337**, 803-812 (2002).
 50. Wallis,G.L., Swift,R.J., Hemming,F.W., Trinci,A.P. & Peberdy,J.F. Glucoamylase overexpression and secretion in *Aspergillus niger*. analysis of glycosylation. *Biochim.Biophys.Acta* **1472**,576-586 (1999).
 51. Takayanagi,T., Kushida,K., Idonuma,K. & Ajisaka,K. Novel N-linked oligo-mannose type oligosaccharides containing an alpha-galactofuranosyl linkage found in alpha-D-galactosidase from *Aspergillus niger*. *Glycoconj.J.***9**,229-234 (1992).
 52. Takayanagi,T., Kimura,A., Chiba,S. & Ajisaka,K. Novel structures of N-linked high-mannose type oligosaccharides containing alpha-D-galactofuranosyl linkages in *Aspergillus niger* alpha-D-glucosidase. *Carbohydr. Res.* **256**,149-158 (1994).

53. Wallis, G.L., Easton, R.L., Jolly, K., Hemming, F.W. & Peberdy, J.F. Galactofuranoic-oligomannose N-linked glycans of alpha-galactosidase A from *Aspergillus niger*. *Eur. J. Biochem.* **268**, 4134-4143 (2001).
54. Wallis, G.L. *et al.* The effect of pH on glucoamylase production, glycosylation and chemostat evolution of *Aspergillus niger*. *Biochim Biophys. Acta* **1527**, 112-122 (2001).
55. Gunnarsson, A., Svensson, B., Nilsson, B. & Svensson, S. Structural studies on the O-glycosidically linked carbohydrate chains of glucoamylase G1 from *Aspergillus niger*. *Eur. J. Biochem.* **145**, 463-467 (1984).
56. Punt, P.J. *et al.* Identification and characterization of a family of secretion-related small GTPase-encoding genes from the filamentous fungus *Aspergillus niger*: a putative SEC4 homolog is not essential for growth. *Mol. Microbiol.* **41**, 513-525 (2001).
57. Oksanen, J., Ahvenainen, J. & Home, S. Microbial cellulases for improving filterability of wort and beer, in Proceedings of the 20th Congress of the European Brewers Convention (IRL Press, Oxford; 1985).
58. Christov, L.P., Szakacs, G. & Balakrishnan, H. Production, partial characterization and use of fungal cellulase-free xylanases in pulp bleaching. *Process Biochem.* **34**, 511-517 (1999).
59. Viikari, L., Kantelinen, A., Sundquist, J. & Linko, M. Xylanases in Bleaching - From an idea to the industry. *FEMS Microbiol. Rev.* **13**, 335-350 (1994).
60. Maat, J. *et al.* Xylanases and their applications in bakery, in *Progress in Biotechnology 7: Xylans and Xylanases* (eds Visser, J. Beldman, G., Kusters-van Someren, M. A. & A.G.J. Voragen, A.G.J.) 349-360 (Elsevier Science, Amsterdam, 1992).
61. Poutanen, K. Enzymes. An important tool in the improvement of the quality of cereal foods. *Trends Food Sci. Technol.* **8**, 300-306 (1997).
62. Bedford, M.R. & Classen, H.L. The influence of dietary xylanase on intestinal viscosity and molecular weight distribution of carbohydrates in rye-fed broiler chicks, in *Progress in Biotechnology 7: Xylans and Xylanases* (eds Visser, J. Beldman, G. Kusters-van Someren, M.A. & Voragen, A.G.J.) 361-370 (Elsevier Science, Amsterdam, 1992).
63. Zeikus, J.G., Lee, C., Lee, Y.E. & Saha, B.C. Thermostable saccharidases. New sources, uses and biodesign. *ACS Symp. Ser.* **460**, 36-51 (1991).
64. Ganter, C., Bock, A., Buckel, P. & Mattes, R. Production of thermostable, recombinant alpha-galactosidase suitable for raffinose elimination from sugar beet syrup. *J. Biotechnol.* **8**, 301-310 (1988).
65. Mulimani, V.H. & Ramalingam, R. Enzymic hydrolysis of raffinose and stachiose in soymilk by alpha-galactosidase from *Gibberella fujikuroi*. *Biochem. Mol. Biol. Int.* **36**, 897-905 (1995).
66. de Vries, R.P. & Visser, J. *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. *Microb. Mol. Biol. Rev.* **65**, 497-522 (2001).
67. de Vries, R.P. Regulation of *Aspergillus* genes encoding plant cell wall polysaccharide

- degrading enzymes; relevance for industrial production. *Appl. Microbiol. Biotechnol.* **61**, 10-20 (2003).
68. Tsukagoshi, N., Kobayashi, T. & Kato, M. Regulation of the amylolytic and (hemi-)cellulolytic genes in aspergilli. *J. Gen. Appl. Microbiol.* **47**, 1-19 (2001).
69. Coutinho, P.M. & Henrissat, B. Carbohydrate-active enzymes: an integrated database approach, in *Rec. Adv. Carbohydr. Bioeng.* (eds Gilbert, H.J. Davies, G.J., Henrissat, B. & Svensson, B.) 3-14 (The Royal Society of Chemistry, Cambridge, 1999).
70. Albersheim, P., Darvill, A.G., O'Neill, M.A., Scols, H.A. & Voragen, A.G.J. An hypothesis: the same six polysaccharides are components of the primary cell walls of all higher plants in *Progress in Biotechnology 14: Pectins and Pectinases* (eds Visser, J. & Voragen, A.G.J.) 47-53 (Elsevier, Amsterdam, 1996).
71. Vincken, J-P. *et al.* Pectin- the hairy thing, in *Adv. in Pectin and Pectinase Research* (eds Voragen, F., Schols, H. & Visser, R.) 47-57 (Kluwer Academic Publishers, Dordrecht, 2003).
72. Vidal, S. *et al.* Structural characterization of the pectic polysaccharide rhamnogalacturonan II : evidence for the backbone location of the aceric acid-containing oligoglycosyl side chain. *Carbohydr. Res.* **326**, 227-294(2000).
73. Christgau, S. *et al.* Pectin methyl esterase from *Aspergillus aculeatus*: expression cloning in yeast and characterization of the recombinant enzyme. *Biochem. J.* **319**, 705-712 (1996).
74. de Vries, R.P., Kester, H.C.M., Poulsen, C.H., Benen, J.A.E. & Visser, J. Synergy between accessory enzymes from *Aspergillus* in the degradation of plant cell wall polysaccharides. *Carbohydr. Res.* **327**, 401-410 (2000).
75. Armand, S. *et al.* The active site topology of *Aspergillus niger* endopolygalacturonase II as studied by site-directed mutagenesis. *J. Biol. Chem.* **275**, 691-696(2000).
76. Pagès, S., Heijne, W.H., Kester, H.C., Visser, J. & Benen, J.A. Subsite mapping of *Aspergillus niger* endopolygalacturonase II by site-directed mutagenesis. *J. Biol. Chem.* **275**, 29348-29353 (2000).
77. Kofod, L.V. *et al.* Cloning and characterization of two structurally and functionally divergent rhamnogalacturonases from *Aspergillus aculeatus*. *J. Biol. Chem.* **269**, 29182-29189 (1994).
78. Suykerbuyk, M.E.G. *et al.* Cloning and characterization of two rhamnogalacturonan hydrolase genes from *Aspergillus niger*. *Appl. Environm. Microbiol.* **63**, 2507-2515(1997).
79. Martens- Uzunova, E.S. *et al.* A new group of exo-acting family 28 glycoside hydrolases of *Aspergillus niger* that are involved in pectin degradation. *Biochem. J.* **400**, 43-52 (2006).
80. Mutter, M., Beldman, G., Pitson, S.M., Schols, H.A. & Voragen, A.G.J. Rhamnogalacturonan α -D-galactopyranosyluronohydrolase. An enzyme that specifically removes the terminal nonreducing galacturonosyl residue in rhamnogalacturonan regions of pectin. *Plant. Physiol.* **117**, 153-163 (1994).

81. de Groot, P.W., Hellingwerf, K.J. & Klis, F.M. Genome-wide identification of fungal GPI proteins. *Yeast* **20**, 781-796 (2003).
82. Adams, D.J. Fungal cell wall chitinases and glucanases. *Microbiol.* **150**, 2029-2035 (2004).
83. Hemker, M. *et al.* Identification, cloning, expression, and characterization of the extracellular acarbose-modifying glycosyltransferase, AcbD, from *Actinoplanes* sp. strain SE50. *J. Bacteriol.* **183**, 4484-4492 (2001).
84. Bobbitt, T.F., Nordin, J.H., Roux, M., Revol, J.F. & Marchessault, R.H. Distribution and conformation of crystalline nigeran in hyphal walls of *Aspergillus niger* and *Aspergillus awamori*. *J. Bacteriol.* **132**, 691-703 (1977).
85. van Leeuwen, C.C., Weusthuis, R.A., Postma, E., van den Broek, P.J. & van Dijken, J.P. Maltose/proton co-transport in *Saccharomyces cerevisiae*. Comparative study with cells and plasma membrane vesicles. *Biochem. J.* **284**, 441-445 (1992).
86. Boze, H., Moulin, G. & Galzy, P. Uptake of galactose and lactose by *Kluyveromyces lactis*: biochemical characteristics and attempted genetical analysis. *J. Gen. Microbiol.* **133**, 15-23 (1987).
87. Osoegawa, K. *et al.* An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* **52**,1-8(1998).
88. Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular Cloning, A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York,1989).
89. Ewing, B., Hillier, L., Wendl, M.C. & Green, P. Base-calling of automated sequencer traces using phred. *Genome Res.* **8**,175-194(1998).
90. Salamov, A.A. & Solovyev, V.V. *Ab initio* gene finding in Drosophila genomic DNA. *Genome Res.* **10**, 516-522(2000).
91. Lukashin, A.V. & Borodovsky, M. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26**, 1107-1115(1998).
92. Burge, C.& Karlin, S. Prediction of complete gene structures in human genomic DNA. *Mol. Biol.* **268**, 78-94(1997).
93. Birney, E., Clamp, M.& Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988-95(2004).
94. 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).
95. Lowe, T.M. & Eddy, S.R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**,955-964(1997).

LEGENDS SUPPLEMENTARY FIGURES

Supplementary Figure 1 Phylogeny of filamentous fungi. Maximum likelihood tree based on concatenation of twenty orthologous proteins ([Supplementary Table 3](#)). Branch length

and likelihood values were calculated by TREE –PUZZLE and indicate substitutions per site. Numbers at nodes are bootstrap values.

Supplementary Figure 2 Cell wall integrity pathway. Signaling proteins of the *S. cerevisiae* Cell Wall Integrity (CWI) pathway and their proposed orthologs in the *A. niger* genome. Cell wall stress is detected by the plasma membrane-localized sensor proteins (*Wsc* proteins). This causes the activation of the small GTPase Rho1p, a step that involves the Rom1/2p GDP/GTP exchange proteins. GTP-bound Rho1p activates the Pkc1-controlled MAP-kinase cascade that is comprised of Bck1p, Mkk1/2p, and Mpk1p. The main target of the CWI pathway is the MADS-box transcription factor Rlm1p that controls the induction of several genes involved in cell wall reinforcement. The consensus sequence for the Rlm1p binding site (RBS; CTA(T/A)4TAG) in the promoter region of cell wall stress-induced genes is conserved between *S. cerevisiae* and *A. niger*.

Supplementary Figure 3 Central metabolism. Glycolysis, hexose monophosphate pathway and glycogen, trehalose and glycerol metabolism in *A. niger*. Gene products contributing to these pathways are indicated. An numbers in italics are for proteins that show significant homology to the established enzyme in the pathway but are not necessarily performing the same reaction. GLYCOLYSIS: GK, glucokinase; HK, hexokinase; PGM, phosphoglucomutase; GPI, Glucose-6-phosphate isomerase; PFK, 6-phosphofructokinase; FBA, Fructose-biphosphate aldolase; TPI, Triose-phosphate isomerase; GAPDH, Glyceraldehyde-3-phosphate dehydrogenase; PGK, Phosphoglycerate kinase; PGAM, Phosphoglycerate mutase; ENO, enolase; PYK, pyruvate kinase. HEXOSE MONOPHOSPHATE PATHWAY: GPDH, Glucose-6-phosphate 1-dehydrogenase; PGL, 6-phosphogluconolactonase; PGD, 6-phosphogluconate dehydrogenase; RPE, Ribulose-5-phosphate 3-epimerase; TK, Transketolase; TA, Transaldolase. GLUCONEOGENESIS: PYCK, Phosphoenolpyruvate carboxykinase; FBP, Fructose-bisphosphatase. ETHANOL PATHWAY: PDC, pyruvate decarboxylase; ADH, alcohol dehydrogenase. GLYCOGEN METABOLISM: GSI, glycogen synthesis initiator; GS, Glycogen synthase; UGP, UDP-glucose pyrophosphorylase; GP, Glycogen phosphorylase. TREHALOSE METABOLISM: TPS, Trehalose 6-phosphate synthase; TPP, Trehalose 6-phosphate phosphatase; NT, neutral trehalase; AT, acid trehalase; AGT, trehalose transporter. GLYCEROL METABOLISM: GFD, Glycerol 3-phosphate dehydrogenase; GPP, Glycerol 3-phosphate phosphatase; DHAP, Dihydroxy acetone phosphate phosphatase; GLD, Glycerol dehydrogenase; DHAK, dihydroxyacetone kinase; GLK, glycerol kinase; GT, glycerol transporter; GF/AQ, glycerol facilitator/aquaporin; DHAS, dihydroxyacetone synthase; MGFD, mitochondrial glycerol 3-phosphate dehydrogenase.

Supplementary Figure 4 G protein and cAMP signaling. Accession numbers are indicated

for all identified components of these signaling pathways. G-protein coupled receptors (GPCRs) are classified according to Lafon *et al.*, 2006. FadA, GanA, GanB, GagC: heterotrimeric G protein alpha subunit. SfaD: heterotrimeric G protein beta subunit. GpgA: heterotrimeric G protein gamma subunit (truncated ORF at 5'end of contig An01c0160). FlbA, RgsA: RGS proteins regulating FadA and GanB, respectively, it is not known whether these proteins may also regulate the GTPase activity of GanA and the *A. niger*-specific GagC heterotrimeric G protein alpha subunits. CyaA: adenylate cyclase. CapA: cyclase associated protein. PdeA: low-affinity phosphodiesterase. PdeB: high affinity phosphodiesterase. PkaR: cAMP-dependent protein kinase regulatory subunit; note that *A. niger* has two genes encoding such regulatory subunits; PkaC, PkaB: cAMP-dependent protein kinase catalytic subunit.

Supplementary Figure 5 Phylogeny of citrate synthases. Phylogenetic analysis of *A. niger* citrate synthases aligned with ClustalX 1.8. Phylogenetic analyses were performed in MEGA 2.1 using the Minimum Evolution model. Stability of clades was evaluated by 1000 bootstrap rearrangements. Bootstrap values lower than 50% and branch length values lower than 0.01 are not displayed in the cladogram.

Supplementary Figure 6 Phylogeny of GH13 glycoside hydrolases. The bootstrapped phylogenetic tree shows the relationship of all GH13-type enzymes identified in the genome sequences of *A. niger*, *A. fumigatus*, *A. nidulans* and *A. oryzae*. Putative localization and function based on sequence similarity are indicated. The alignment and phylogenetic analysis were performed with MEGA version 3.1 using default settings. A bootstrapped tree was constructed with the neighbor-joining method using 500 replicates.

Supplementary Figure 7 Putative *A. niger* fumonisin cluster. Orthologs of the *Gibberella moniliformis* fumonisin gene cluster are connected to the gene cluster region of *A. niger*. Genes at the left and right hand borders of the *A. niger* cluster region do not appear to be secondary metabolism genes. *G. moniliformis* genes which have been shown by deletion and/or over-expression to be involved in fumonisin biosynthesis are highlighted by cross-hatching. Deletion of *fum19* (ABC transporter) had a slight effect, and the *fum19* ortholog in *A. niger* is the only gene for which significant transcription could be detected under standard fermentation conditions (highlighted in black). Orthologs of *Gibberella fum11* (tricarboxylate transporter), *fum12* (P450 hydroxylase) and *fum16* (fatty acyl-CoA synthetase) were not identified in the *A. niger* genome. Predicted gene functions are: *fum1* polyketide synthase, *fum6* cytochrome P450 monooxygenase, *fum7* dehydrogenase, *fum8* aminotransferase, *fum9* dioxygenase, *fum10* fatty acyl-CoA synthetase, *fum11* tricarboxylate transporter, *fum12* cytochrome P450 monooxygenase, *fum13* short chain dehydrogenase, *fum14* NRPS-like condensation domain, *fum15* cytochrome P450 monooxygenase, *fum16* fatty acyl-CoA

synthetase, *fum17* longevity assurance factor, *fum18* longevity assurance factor, *fum19* ABC transporter.