Yu-Fei Sui, Li-Ming Ouyang, Tabea Schütze, Shu Cheng, Vera Meyer, Ying-Ping Zhuang

Comparative genomics of the aconidial Aspergillus niger strain LDM3 predicts genes associated with its high protein secretion capacity

WISSEN IM ZENTRUM
UNIVERSITÄTSBIBLIOTHEK

Technische
Universität
Berlin

Comparative genomics of the aconidial *Aspergillus niger* strain LDM3 predicts genes associated with its high protein secretion capacity

Yu-Fei Sui[1, 2,†], Li-Ming Ouyang[1,†], Tabea Schütze[2], Shu Cheng[3], Vera Meyer[2,*], Ying-Ping Zhuang[1, *]

1 State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, P. R. China;

2 Department of Applied and Molecular Microbiology, Institute of Biotechnology, Technische Universität Berlin, Gustav-Meyer-Allee 25, 13355 Berlin, Germany

3 BGI Institute of Applied Agriculture, BGI-Shenzhen, Shenzhen 518120, CHINA

† Both authors contributed equally.

* Corresponding authors

Correspondence: Vera Meyer (vera.meyer@tu-berlin.de); Yingping Zhuang (ypzhuang@ecust.edu.cn)

Abstract

Aspergillus niger is widely used as a cell factory for homologous and heterologous protein production. As previous studies reported that reduced sporulation favors protein secretion in *A. niger*, in this study we conducted a comparative genomic analysis of the non-sporulating industrially exploited *A. niger* strain LDM3 in China and the model protein secretion strain CBS 513.88 to predict the key genes that might define the genetic basis of LDM3's high protein producing potential in silico. After sequencing using a hybrid approach combining Illumina and PacBio sequencing platforms, a high-quality genome sequence of LDM3 was obtained which harbors 11,209 open reading frames (ORFs) and exhibits large chromosomal rearrangements in comparison to CBS 513.88. An alignment of the two genome sequences revealed that the majority of the 457 ORFs uniquely present in LDM3 possessed predicted functions in redox pathways, protein transport, and protein modification processes. In addition, bioinformatic analyses revealed the presence of 656 ORFs in LDM3 with non-synonymous mutations encoding for proteins related to protein translation, protein modification, protein secretion, metabolism and energy production. We studied the impact of two of these on protein production in the established lab model strain N402. Both *tupA* and *prpA* genes were selected because available literature suggested their involvement in asexual sporulation of *A. niger*. Our co-expression network analysis supportively predicted the role of *tupA* in protein secretion and the role of *prpA* in energy generation, respectively. By knockout experiments, we showed that the Δ*tupA* mutant displayed reduced sporulation (35%) accompanied by higher total protein secretion (65%) compared to its parental strain. Such an effect was, however, not observed in the Δ*prpA* mutant.

Keywords: Comparative genomics; *Aspergillus niger*; sporulation; protein secretion; *tupA*; gene co-expression network

Introduction

Aspergillus niger is a widely used cell factory in bulk manufacturing of industrial enzymes and organic acids (Cairns et al. 2018, 2019). To date, fungi have been a crucial source of the majority of antibiotics (Liu et al. 2010, 2012). As it has recently been reprogrammed to produce secondary metabolites and pharmaceutical ingredients at a high level, *A. niger* is of general interest as a multipurpose cell factory (Boecker et al. 2018). Presently, a total of 17 *A. niger* genomes have been sequenced since the first *A. niger* genome became available in 2007 (Andersen et al. 2011; Baker 2006; Gong et al. 2016; Paul et al. 2017; Pel et al. 2007; Vesth et al. 2018; Yin et al. 2014; Yin et al. 2017). Remarkably, all *A. niger* strains have high genome flexibility and share about 7,500 genes in their core genome but differ in hundreds up to thousands of genes, which define the pan-genome and species-unique genes, respectively (Vesth et al. 2018).

A. niger LDM3 is an industrial glucoamylase (GlaA) production strain which features a very high GlaA production level and is phenotypically characterized by an aconidial phenotype. This is of special interest because another high efficient GlaA-producing *A. niger* strain exploited in China (strain SH2) is non-sporulating as well (Yin et al. 2014). Most interestingly, solid-state fermentations uncovered that proteins are mainly secreted in the central and peripheral regions of *A. niger* colony but not in the mycelial regions undergoing sporulation, indicating that sporulation inhibits protein secretion (Krijgsheld et al. 2013). Swift et al. (1998) have already proved about 20 years ago that the aconidial phenotype of *A. niger* is beneficial to protein biosynthesis and/or secretion. Several mutants with reduced sporulation were isolated from maltodextrin-limited chemostat and pH auxostat cultivations of *A. niger* strain B1 (carrying 20 copies of the GlaA encoding gene *glaA*). Two of these spontaneous morphological mutants showed almost yellow and white colonies when cultivated on agar plates and presented a significantly improved GlaA production compared to their parental strain B1, even though one of the mutants showed that more than half of the *glaA* copies were lost. Similarly, Jorgensen et al. (2011a) obtained two sporulation deficient *A. niger* strains scl-1 and scl-2 through UV-mutagenesis, in which several secondary metabolites were produced less but secreted proteins were remarkably accumulated. However, the molecular mechanisms linking protein secretion and asexual sporulation are not fully understood so far. In general, sporulation-deficient *Aspergillus* strains are known to be defective in many regulators, including the transcription factors (TFs) BrlA and FlbA. The

functions of both these 2 proteins have been well documented in *A. niger*, *Aspergillus nidulans*, *Aspergillus fumigatus*, and *Aspergillus oryzae* (Adams et al. 1988; Krijgsheld et al. 2013; Lee and Adams 1996; Mah and Yu 2006; Pavezzi et al. 2011; van Munster et al. 2015; Yamada et al. 1999). BrlA is the central regulator of conidiophore development which becomes activated by FlbA (Krijgsheld et al. 2013). Notably, the deletion of *flbA* gene in *A. niger* results in a fluffy phenotype, accompanied by a thinner cell wall and a more complex secretome (Krijgsheld et al. 2013).

To shed light on the molecular mechanisms behind the aconidial and high-secretion phenotype of LDM3 strain, we sequenced its genome by a hybrid approach combining PacBio RS and Illumina HiSeq 4000 technologies and compared this genome to the GlaA producing model strain CBS 513.88. In addition, gene knockout experiments were performed with two genes of our interest, *tupA* and *prpA*, to investigate their impact on protein secretion.

Materials and Methods

Strains and culture

A. niger strains used in this study are listed in Table 1. The *A. niger* strain LDM3 with aconidial phenotype was kindly provided by Longda Biotechnology (Shandong, China). Czapek–Dox slope and submerged medium were used to cultivate LDM3. The composition of Czapek–Dox slope medium is as follows: sucrose 3%, NaNO$_3$ 0.2%, MgSO$_4$•7H$_2$O 0.05%, KCl 0.05%, FeSO$_4$•7H$_2$O 0.001%, K$_2$HPO$_4$ 0.1%, and agar 1.7%, pH 5.5~6.0. LDMS was cultivated at 34°C for 5 days. The composition of the Czapek–Dox submerged medium is the same as the slope medium specified above, but without agar. Cultivation was performed at 34°C, 180~200 rpm for 72 h and the pellets were collected by filtration.

The other *A. niger* strains were cultured as follows: Strains were grown at 30 °C using the complete or minimal medium (Arentshorst et al. 2012) and supplemented with 1 mM uridine where necessary. To test the yield of GlaA among different mutants, 10^6 spores/mL of strains FW35.1 (Wanka et al. 2016), YS33.10 (*kusA*::DR-*amdS*-DR, *tupA*::*AopyrG*, *pyrG+*), and YS34.16 (*kusA::DR-amdS-DR, prpA::pyrG, pyrG*$^+$) were inoculated into 50 mL CM liquid medium with 3% w/v glucose as the carbon source and cultivated at 30°C and 250 rpm. Samples were taken at 24, 48, 72, 96, and 120 h after inoculation. Physiological parameters

(dry weight, total secreted protein, residual glucose concentration and enzyme activity of GlaA in the media) were measured. Experiments were performed in biological quadruplicates.

Genome DNA extraction and sequencing

A. niger transformation, genomic DNA extraction, and Southern hybridization were performed as previously described (Arentshorst et al. 2012). Quality analysis of genomic DNA, library construction, and sequencing on PacBio (RS II) and Illumina (HiSeq 4000) instruments were performed by BGI (Shenzhen, China).

Hybrid assembly of the *A. niger* LDM3 genome sequence using Illumina and PacBio sequencing

The development of several hybrid genome assembly algorithms allows taking of reads from multiple read sources (Chen et al. 2017; Rhoads and Au 2015). Reads from the Illumina platform are short but accurate, while reads from the PacBio are long but accompanied by a high error rate. Hence, hybrid sequencing allows the use of long reads for genome assembly (PacBio reads) while Illumina reads can be used for corrections. A total of 296,149 subreads (2.37 Gb) were generated on the PacBio RS II platform with an average length of 8,006 bp, and 6.45 Gb of clean data (read length 150 bp, insert size 300 bp) were generated on the Illumina HiSeq 4000 platform. The subreads were self-corrected and then assembled using Falcon v 0.3.0 (https://github.com/PacificBiosciences/falco). The resulting assembly was corrected through Illumina reads using Proofread v 2.12 (https://github.com/BioInf-Wuerzburg/proovread).

Bioinformatic analyses

Gene models were predicted using Augustus v 3.2.1 (http://bioinf.uni-greifswald.de/augustus/), SNAP v 2010-07-28 (http://korflab.ucdavis.edu/software.html) and GeneMark-ES v 4.28 (http://exon.gatech.edu/) with *A. niger* CBS 513.88 as the reference. Gene structure was predicted using GeneWise v 2.20 (http://www.sanger.ac.uk/Software/Wise2/). The predicted gene models were functionally annotated by aligning their protein sequences against the KEGG (Minoru et al. 2015), SwissProt (http://www.gpmaw.com/html/swiss-prot.html), GO (Ashburner et al. 2000), COG (Clusters of Orthologous Groups) (Galperin et al. 2014), KOG (EuKaryotic Orthologous Groups) (Huerta-Cepas et al. 2015), TrEMBL (https://www.uniprot.org/), and Non-Redundant protein databases (https://www.ncbi.nlm.nih.gov/) with BLASTP (E-value $\leqq$ 1.0e-5). tRNAScan-SE v 1.3.1 (http://gtrnadb.ucsc.edu/) was used for tRNA prediction.

SNP, InDel and SV analyses

High quality filtered short reads from the Illumina platform were mapped to the reference genome via BWA (Burrows-Wheeler Aligner) (Li and Durbin 2010). After filtering for Q < 20, the paired-end (PE) reads were aligned to all chromosomes and the aligned PE reads with a distance of > 1000 bp were screened for the genome assembly. GATK v1.6-13 (http://www.broadinstitute.org/gatk/) was used to detect single nucleotide polymorphism (SNP), insertion and deletion (InDel) between *A. niger* LDM3 and the reference CBS 513.88 based on high-quality alignment results. Raw SNPs and InDels were filtered under a stringent criterion of GATK Unified Genotyper (Yin et al. 2014).

Synteny analysis

The assembled *A. niger* LDM3 genome sequence was fragmented into 1 kb length and was compared to the genome sequence of *A. niger* CBS 513.88 by BLASTN using the cutoff value 1×10^{-75}. The sequence of the target fungus was sorted according to that of the reference fungus based on MUMmer alignment results (Kurtz et al. 2004). Synteny analysis was performed as previously described (Andersen et al. 2011).

Strain-specific genes in the LDM3 genome compared to CBS 513.88

The imprint algorithm used to determine strain-specific genes in LDM3 was performed as previously described (Andersen et al. 2011). Using BLASTP (E-value ≤ 1.0e-10, identity > 50%, coverage > 80%) the CDSs of *A. niger* LDM3 was compared to the amino acids sequence of *A. niger* CBS 513.88. The corresponding amino acids sequence from CBS 513.88 was named as 'Imprint'. For the comparison of CDSs and Imprint, gene variations such as InDel, synonymous mutations, frameshift mutations, and partial hit for each CDS pair were collected into gene mutation lists. To extract the corresponding full-length nucleotide sequence of the genes that were not aligned to CBS 513.88 under the above BLASTP criteria, genes with 100% alignment to the reference were removed. In addition, the remaining genes were manually analyzed for frameshift mutation, start codon and stop codon loss, partial match (coverage < 50%), early termination and no hits in LDM3 compared to CBS 513.88.

Co-expression network of TupA and PrpA encoding genes

From 283 microarray experiments of *A. niger* hosted by FungiDB (Basenko et al. 2018), the co-expression networks of TupA and PrpA encoding genes were retrieved according to a previous study (Schape et al. 2019). Gene pairs passing a Spearman's correlation coefficient of |0.7| were used to construct co-expression networks. For the TupA network, 32 genes were negatively correlated, while for PrpA, this number increased to 576 genes. Both TupA and PrpA networks were assessed for GO-enriched biological process relative to *A. niger* genome using default parameters in FungiDB, and the genes of interest were manually filtered when Benjamini–Hochberg FDR corrected *p*-values were above 0.05.

Determination of biomass dry weight, total secreted protein, residual glucose and enzyme activity of GlaA

 4 ml broth was taken at the indicated time points from shake flask cultures. Biomass and culture supernatant were separated by vacuum filtration followed by 3 times washing with deionized water, frozen at − 80 °C, and freeze-dried overnight for the determination of biomass. Total extracellular protein in the culture supernatant was determined via the Bradford assay (BioRad, Hercules, CA, USA) according to the manufacturers' protocols, and absorbance (600 nm) was measured using the GloMax®-Multi Detection System (Promega, Madison, USA). Quantification of residual glucose in the cultivation medium was performed using the Glucose GOD/PAP kit (Human, Wiesbaden, Germany) according to the manufacturer's manual. Enzyme activity is expressed in AGI units, which is related to an officially assigned GlaA standard. One AGI unit is defined as the amount of enzyme that produces 1 µmol glucose per min at 60 °C and pH 4.3 from the soluble starch substrate. 20 µl supernatant was mixed with 230 µl *p*-NPG substrate (2 g/l 4-nitrophenyl α-D-glucopyranoside acetate buffer pH 4.3, pre-warmed for 5 min at 37 °C). After incubation at 37 °C for 20 min, 100 µl of 0.3 M Na_2CO_3 was added to stop the reaction and the absorbance was immediately measured at 405 nm using a plate reader. The standard GlaA from *A. niger* (E.C 3.2.1.2; Sigma Aldrich, Darmstadt, Germany) was used to build a standard curve with GlaA enzyme activity$=\dfrac{OD405+0.01}{0.008}$ ×dilution rate ($R^2 > 0.999$).

Generation of Δ*tupA* and Δ*prpA* deletion strains

To improve the homologous recombination efficiency, the split marker method and the non-homologous end joining (NHEJ) deficient recipient strain MA169.4 (Carvalho et al. 2010) were exploited. For gene deletions, deletion cassettes containing homologous 5′ or 3′ flanks (~1.5 kb) for targeted integration and the selective marker *AopyrG* (*A. oryzae*) were constructed. These were co-transformed into the *pyrG-* recipient strain MA169.4, and the transformants were screened based on uridine prototrophy. *A. niger* transformations were carried out using the protoplast transformation method as described in Arentshorst et al. (2012). The 5′ and 3′ flanks were amplified by PCR with the primers described in Supplementary Table S1 and S2. The details about the construction of deletion cassettes were illustrated in Supplementary Fig. S1 and S2, and positive *tupA* or *prpA* deletion strains were confirmed through diagnostic PCR and Southern analysis (Supplementary Fig. S3, S4). The autonomously replicating plasmid pMA171 (Carvalho et al. 2010) was exploited in the complementation studies. The ORF of *tupA* including approximately 0.6 kb promoter and 0.6 kb terminator regions was amplified taking N402 genomic DNA as template and cloned into *Not*I-linearised pMA171 (Supplementary Fig. S5). Then the constructed plasmid pMA171-tupA was transformed into the Δ*tupA* deletion mutant. Primary transformants containing the complementation plasmid were purified on MM medium containing 100 µg/ml of hygromycin and further analyzed by diagnostic PCR.

Data access

The complete chromosomal sequence of LDM3 is available at the GenBank under the assigned accession number VTFN00000000.

Results

Characteristics of *A. niger* LDM3 genome

In order to identify the genetic determinants responsible for the unique phenotype of LDM3, the entire genome of LDM3 was sequenced using a hybrid approach that combined Pacific Biosciences with Illumina sequencing, obtaining 6,447 Mb and 2,679 Mb data after filtering from the Illumina HiSeq 4,000 and PacBio RS II platform respectively. The high-quality reads were further used to assemble the genome of LDM3 after quality control, resulting in a 35.28 Mb genome sequence with 11 scaffolds and a sequencing depth of 177× (Table 2). The assembled genome base calls were corrected with Illumina high-quality PE reads. A total of 11,209 ORFs were identified in LDM3 (94% of the genes were automatically annotated based on protein

databases with an average gene length of 1,691 bp), displaying a lower gene density (0.32 gene/kb) compared to CBS 513.88 (0.42 gene/kb) (Andersen et al. 2011) (Supplementary Table S3).

KOG analysis was employed to identify their biological roles. Out of the 11,209 predicted proteins, 9,712 ORFs (87%) were assigned to 24 KOG functional categories in total (Supplementary Fig. S6). In LDM3, slightly more genes were allocated to term G (Carbohydrate transport and metabolism), A (RNA processing and modification), L (Replication, recombination and repair), O (Posttranslational modification, protein turnover, chaperones) and to the unknown functional gene class (S) compared to the control. Both LDM3 and CBS 513.88 showed a comparable number of predicted genes distributed in each term, defining a high genome similarity. 264 tRNA genes were identified in LDM3, which were comparable to other enzyme-producing strains CBS 513.88 (269) and SH2 (267), but more than other *Aspergillus* strains including *A. nidulans* (188) and *A. fumigatus* (179) (Yin et al. 2014).

Genome structure variation analysis

Strains LDM3 and CSB 513.88 share widely distributed syntenic blocks, accounting for 96.56% of their genomes. The dot plot depicted in Supplementary Fig. S7 shows conserved synteny between the two strains, reflecting a close phylogenetic relationship. However, the synteny map illustrated remarkable chromosomal rearrangements for the 8 chromosomes (19 supercontigs) of CBS 513.88 (Fig. 1). The LDM3 genome was assembled into 11 scaffolds, resembling a higher quality of assembly than the 19 supercontigs of CBS 513.88. The entire lengths of 4 out of 9 scaffolds in LDM3 reached a longer or similar length compared to their corresponding chromosomes in CBS 513.88. Scaffolds 1 and 2 are larger than any chromosomes of the reference strain (6.0 Mb in maximum) reaching 7.6 Mb and 7.5 Mb, respectively. This suggests a fusion with other chromosomes, representing a noticeable structural variation in LDM3. In addition, scaffold 7 and 8 compose only the third supercontig in CBS 513.88.

Interestingly, in general, 2 GlaA encoding genes can be identified in most published *A. niger* genomes, namely *glaA* (An03g06550) and *glaB* (An12g03070, sharing 25% identity with *glaA*), in which *glaA* being more strongly expressed than *glaB* (Andersen et al. 2011; Schape et al. 2019). This is also the case in CBS513.88 and SH2, where only single copy of *glaA* and *glaB* are present in their genomes. However, only a single copy of *glaB* but no *glaA* is present in the LDM3 genome, and is found without any mutants compared to that of the control.

LDM3 strain-specific genes analysis

An alignment of the two genome sequences of LDM3 and CBS513.88 showed some unique regions in LDM3, incorporating 457 protein-encoding genes (Supplementary Table S4). These include 196 ORFs with frameshift mutations, 17 ORFs in which the start or stop codon was lost, 81 ORFS with partial match (coverage < 50%), 4 ORFs with early termination and 157 ORFs which did not match the CBS 513.88 sequence (75 were annotated as hypothetical or of unknown function). To further characterize these strain-specific genes in LDM3, we compared their sequences with publicly available genomes of other *A. niger* and *Aspergillus* strains, demonstrating that orthologs of 49 out of 457 strain-specific genes were identified in all the compared genomes (Supplementary Table S5). GO annotations for 225 out of 457 ORFs are available, which are enriched in catalytic activity, oxidoreductase activity, hydrolase activity, transferase activity, protein binding, oxidation-reduction process, transporter activity, and localization GO terms (Supplementary Fig. S8). Among the genes annotated in 2 significantly enriched GO terms oxidoreductase activity (36 genes) and oxidation-reduction process (30), 7 genes could not be aligned to the reference genome, 7 genes showed frameshift mutations, and 1 gene was mapped partially to CBS 513.88 genome (Supplementary Table S6). The 7 unaligned genes were mainly predicted to function in amino acid metabolic pathways, including degradation of proline, isoleucine and leucine (AN2_GLEAN_10000163, *mmsB* predicted as 3-hydroxybutyrate hydrogenase), biosynthesis of alanine, aspartic acid, and glutamic acid (AN2_GLEAN_10000741, *gabD* predicted as succinate-semialdehyde dehydrogenase which supplements succinic acid for the TCA cycle (Yin et al. 2017), and the biosynthesis of arginine and proline (AN2_GLEAN_10007014, *proA* predicted as glutamate-5-semialdehyde dehydrogenase).

SNP and InDel analysis

Compared to those in the reference genome of CBS 513.88, a total of 2,138 SNP and InDel mutations are present in the genome of LDM3, in which non-synonymous mutations are distributed in 656 ORFs (Supplementary Table S7). KOG clustering analysis uncovered that the mutated genes are mainly clustered in A (RNA processing and modification), C (Energy production and conversion), E (Amino acid transportation and metabolism), G (Carbohydrate transport and metabolism), O (Posttranslational modification, protein turnover, chaperones), J (Translation, ribosomal structure and biogenesis) and I (Lipid

transport and metabolism) (Fig. 2), all of which are very fundamental for a high level of protein expression, protein targeting, and secretion. Selected genes of interest are depicted in Table 3 and will be discussed in detail in the next section.

In vivo analysis of two selected genes putatively involved in sporulation

Given the possible link between the aconidial and high enzyme production phenotype in *A. niger*, we decided to study the function of two genes of our interest (*tupA*, An15g00140 and *prpA*, An18g01170) among the 656 mutated genes in LDM3 compared to CBS 513.88. In doing so, we selected the lab strain MA169.4 as the parental strain, which is devoid of the NHEJ pathway and thus ensures a higher homologous recombination rate (Carvalho et al. 2010). TupA (non-synonymous) is a globally active transcriptional repressor (ortholog of the repressor Rco-1of *Neurospora crassa* and Tup1p of *S. cerevisiae*). Notably, its deletion has been shown to cause an aconidial phenotype and a reduced growth rate in *N. crassa* (Yamashiroy et al. 1996) and *A. niger* (Schachtschabel et al. 2013). Thus, the present mutations in TupA prompted us to investigate the function of *tupA* in the unique phenotype of LDM3. *PrpA,* a gene of unknown function, is absent in LDM3. Its expression is induced by *brlA* and *abaA*-dependent regulatory loops in *A. niger* and is predicted to cause the aconidial phenotype in the *A. niger* SH2 strain (Yin et al. 2014). Blast result indicated that the protein sequence of TupA carries a 16 amino acids insertion and one amino acid change from glycine (G) to aspartic acid (D) (Supplementary Fig. S9).

To predict the function of *tupA* and *prpA* in *A. niger*, we harnessed our recently published genome wide co-expression database available on FungiDB to construct co-expression networks for *tupA* and *prpA* (Schape et al. 2019). The co-expression network of *tupA* showed an exclusively negative correlation with the genes predicted to function in protein secretion, filamentous growth, vesicle-mediated transport, cellular protein metabolic process, and fungal-type cell wall organization or biogenesis (Fig. 3a), and this is consistent with its function as a transcriptional repressor. Similarly, *prpA* showed nearly no positive correlations with other genes but surprisingly a negative correlation with a high number of genes (567). GO enrichment analysis revealed that the *prpA* network is mainly enriched in mitochondrion organization, protein targeting to the mitochondrion and protein catabolic processes, namely associated with energy generation (Fig. 3b). Notably, whereas the *tupA* network contained overrepresented GO terms associated with growth or sporulation, this was not the case for the *prpA* co-expression network (Supplementary Table S8, S9).

To study the impact of both genes on sporulation and protein secretion in *A. niger*, single gene knockout strains were constructed in MA169.4 using the split marker approach as published previously (Fiedler et al. 2018b). As depicted in Fig. 4 and Fig. 5, the deletion of *tupA* severely reduced the mycelial growth rate and sporulation efficiency of *A. niger*, which was not the case for the *prpA* null mutant. All Δ*tupA*-complemented transformants obtained grew like the wild-type, confirming that the severe growth and sporulation defect of mutant was caused by the *tupA* deletion and the plasmid-based *tupA* gene is capable to nearly restore the phenotype (Fig. 4). However, the Δ*tupA*-complemented strain was still witnessed a weaker growth rate, suggesting that the cellular amount of *tupA* is under stringent control. The sporulation capacity of Δ*tupA* was reduced by about 35% compared to the reference strain FW35.1, but Δ*prpA* presented nearly no difference compared to the reference strain (Fig. 5). It has thus been confirmed that knocking out *tupA* in *A. niger* strongly inhibits the growth rate of the strain, which is consistent with a previous report (Schachtschabel et al. 2013).

In order to determine whether the sporulation defect in the *tupA* null mutant could improve the protein production capacity of *A. niger*, the reference strain FW35.1, Δ*prpA,* and Δ*tupA* were cultured at the shake flask level. Pairwise comparison of values for Δ*prpA* and parental strain gave comparable results in protein secretion. While compared to the reference strain FW35.1, the growth rate of Δ*tupA* was slower in the early stage of fermentation which was consistent with the findings of a previous report (Schachtschabel et al. 2013), but dramatically increased after 24 h (Fig. 6b). During the flask-level fermentation, the reference strain showed almost no color change, but the broth color of the Δ*tupA* mutant suddenly turned yellowish after the fourth day and brown on the fifth day (Fig. 6a). Hence, during the submerged fermentation process, there is a variation in growth and physiological characteristics between the mutant strain and the reference.

As it can be seen in the Δ*tupA* fermentation results, in the first 3 days (exponential and early stationary phase of fermentation), Δ*tupA* secreted much less protein especially GlaA which was indeed not detectable compared to the parental strain (Fig. 6c, d). However, after prolonged cultivation (day 4-5), the amount of extracellular protein was considerably accumulated in the Δ*tupA* strain and was significantly higher than that in the control (increased by about 67% at day 5). In particular, the enzyme activity of GlaA at 70 h post-inoculation showed up to 54-fold increase compared to that at 45 h, therefore the dramatic accumulation of GlaA contributed mostly to the increase of extracellular secreted protein.

Discussions

In light of the dependence of superior protein capacity on defective sporulation, we reported in this study a high-quality assembled genome of an industrially relevant *A. niger* strain LDM3 used for GlaA production. This was achieved by a hybrid sequencing approach which utilized Illumina and PacBio sequencing technologies. Owing to its unique phenotype and high protein producing potential, it is of great interest to both theoretical research and biotech industry to explore the novel functional properties of LDM3. Gene functional annotation (Supplementary Fig. S10) confirmed that the majority of genes were significantly allocated in catalysis, transport, translation, carbohydrate metabolism, and amino acid metabolism, which matched well with the high-yield protein-producing characteristic of LDM3. In addition, an increasing number of tRNA genes might suggest that *A. niger* could rely on a higher translation efficiency in comparison to *A. nidulans* and *A. fumigatus*.

Interestingly, *glaB* is witnessed as the only single copy of the GlaA encoding gene in LDM3, albeit *glaA* behaves dominatingly in the majority of characterized *Aspergillus* genomes while being absent in LDM3. Previously reported that *glaB* displayed diverse expression patterns in *A. oryzae* under distinct cultivation conditions, and was strikingly expressed in solid-state cultures but showed little or no expression in submerged cultivation (te Biesebeke et al. 2005). *glaB* is regulated at the transcriptional level, which could be enhanced by starch, low water activity, high temperature and limited hyphal extension (Kumar and Satyanarayana 2009). Similarly, it was also induced by isomaltose in *A. nidulans* (Nakamura et al. 2006). Given the abundant expression profiles of *glaB,* its performance in LDM3 requires further description in submerged cultures.

LDM3 and CBS 513.88 exhibited a close phylogenetic relationship, sharing 97% identity. However, the unique regions between the two genome sequences mainly define the genome diversity. Among the 872 TFs predicted in *A. niger* (Park et al. 2008; Szklarczyk et al. 2017), 9 TFs thereof carry InDel mutations in LDM3 (Supplementary Table S10). Among those is CpcA, which is involved in the degradation of misfolded proteins and plays a role in RNA processing and translation processes by the endoplasmic reticulum protein response pathway (Jorgensen et al. 2009; Wanke et al. 1997). CpcA is the functional ortholog of the *Saccharomyces cerevisiae* transcriptional activator Gcn4p in *Aspergillus* and induces the expression of multiple genes associated with amino acid biosynthesis under amino acid starvation conditions

(Vongsangnak et al. 2011). Previous reports indeed assumed that a frame shifted *cpcA* gene might be the cause of efficient expression of the GlaA encoding gene *glaA* and other enzymes in *A. niger* (Andersen et al. 2011; Yin et al. 2014). It is thus very tempting to speculate that this is also the case with LDM3.

Moreover, 41 out of 60 TFs containing SNP display non-synonymous mutations (Supplementary Table S10). A wide range of regulatory pathways, including starch degradation, cell wall synthesis, nitrogen assimilation, and amino acid synthesis pathways were affected. AmyR (An04g06910), for example, is an essential maltose-dependent TF that regulates the expression of starch hydrolase genes such as extracellular hydrolases including alpha-amylase AamA, alpha-glucosidases AgdA and AgdB, and GlaA. The consensus sequence of DNA binding of AmyR to the promoter regions of its target genes is well known (CGGN8(C/A)GG) (Yuan et al. 2008). Such a sequence can indeed be found at position -878 bp upstream of *glaB,* suggesting that the *glaB* gene in LDM3 might be under transcriptional control of AmyR as in other *Aspergillus*.

The majority of transporters carrying SNP mutations belong to the MFS family of secondary active transporters and facilitators (de Vries et al. 2017), such as glucose transporters An15g04270 and An11g09600 which may be beneficial for the uptake of substrates in LDM3. In addition, there are 3 neutral amino acid transporters An07g03690, An07g03970, and An14g07130 distributed non-synonymous mutations. Ala, Leu, Thr, and Ser are the top 4 amino acids of GlaA, and interestingly, all of them are neutral amino acids (Supplementary Table S11). It can thus be posited that the variation of neutral amino acid transporters could be relevant to the transport of these 4 primary amino acids, to support the translation of *glaB* mRNA.

Within the 465 genes predicted to function in the secretory pathways of *A. niger* (Carvalho et al. 2012; Guillemette et al. 2007; Jorgensen et al. 2009; Kwon et al. 2012), 3 take InDel mutations and 25 contain non-synonymous SNPs in LDM3 (Supplementary Table S12). Most genes are involved in protein transport, unfolded protein response, glycosylation, and starch metabolism, or are protease encoding genes. Two genes with non-synonymous mutations are worth mentioning: An07g02190 (*S. cerevisiae sec7* ortholog) plays a predicted role in vesicles traffic in intra-Golgi and ER-to-Golgi transport (Wolf et al. 1998) and An08g00290 (*S. cerevisiae rud3* ortholog) is a matrix protein of the Golgi and essential for its integrity (Gillingham et al. 2004).

Proteases are capable of hydrolyzing protein peptide chains, which may be detrimental to the accumulation of secreted protein. The 3 mutated protease genes (*pepC, pepF* and *cpy1p*) in LDM3 are all serine proteases.

Since the dominating status of serine and threonine in the composition of GlaA (Supplementary Table S11), the reduced degradation of serine enriched proteins caused by the mutations of the serine protease genes has aroused the interest to merit further exploration in the future.

The fungal cell wall determines the hyphal morphology, cellular integrity, and protein secretion productivity during the growth and development of *A. niger* (Cairns et al. 2019). Variations in cell wall composition and mycelial morphology of *A. niger* can thus support protein production (Fiedler et al. 2018a, b). For example, *gelB* (An08g07350) encoding a GPI-anchored glucosyltransferase important for β-1,3-glucan synthesis, is affected (non-synonymous) (Supplementary Table S13). Deletion of *gelB* in *A. fumigatus* causes the reduction of β-1,3-glucan cell wall levels accompanied by abnormal germination, decreased growth and deficient pigment biosynthesis during sporulation (Mouyna et al. 2005). However, it has not been reported so far whether mutations on the glucan biosynthesis pathway lead to the aconidial phenotype of *Aspergillus*. Another interesting mutated gene is a chitin synthase encoding gene An02g02360 (*csmA*) with non-synonymous SNPs in LDM3. Orthologs are known for *A. fumigatus* (*chsE*) and *A. nidulans (csmA)* and their deletion phenotypes have been studied. Deletion of the *A. fumigatus chsE* gene provokes abnormalities in hyphal morphology, sporulation, and reduced spore survival rate (Aufauvre-Brown et al. 1997; Jimenez-Ortigosa et al. 2012), and deletion of the *A. nidulans csmA* induces balloon-like swollen hyphae and intrahyphal hyphae formation (Takeshita et al. 2002, 2005). These observations indicated that chitin is essential for maintaining conidiophore and spore integrity (Jimenez-Ortigosa et al. 2012). However, other chitin synthase gene knockout mutants (*chsA, chsB, chsC, chsD, chsG,* and *chsF*) did not introduce abnormal spore formation in *A. fumigatus* (Mellado et al. 1996a, b, 2003; Rogg et al. 2011). Interestingly, in addition to the aconidial phenotype, LDM3 also underwent a unique morphological transition phase during the late stages of submerged bioreactor cultivations, which is not common to other *A. niger* strains (Supplementary Fig. S11). In fed-batch cultures, LDM3 mycelia began to swell at their tips during oxygen limitation, followed by mycelial fragmentation and separation of dispersed mycelial structures into smaller entities (data not shown). This finding assumed that this morphological response improved oxygen transfer. Notably, GlaA dramatically accumulated in the supernatant, suggesting that hyphal swelling and fragmentation are important for high GlaA production or release into the medium (data not shown). It is thus tempting to

speculate that the mutation of *csmA* may be relevant to the morphological adaptation of LDM3 during fed-batch fermentation.

It is well known that most species of *Aspergillius* reproduce asexually. A central regulatory pathway (*brlA* → *abaA* → *wetA*) is conserved in all *Aspergillus* and *Penicillium* genomes and controls conidial-specific gene expression and asexual sporulation (de Vries et al. 2017). BrlA is required to activate *abaA* and *wetA* (Yin et al. 2014). *FlbA* is the central regulator, controlling the binding of FlbB and BrlA to the G-protein coupled receptor, which represses the growth of vegetative mycelium. Typically, a deletion in *flbA* in filamentous fungi induces abnormal conidiation (Lee and Adams 1994; Perez-de-Nanclares-Arregi and Etxebeste 2014; van Munster et al. 2015; Wieser et al. 1994). Moreover, double deletion of *brlA* and *flbA* results in the fluffy phenotype of *A. niger* (van Munster et al. 2015). In LDM3, only synonymous mutations were distributed in the respective *flbA* homolog and no mutation in *brlA* or *flbB* was identified, therefore these 3 genes are likely not related to the aconidial phenotype of LDM3. An12g02050 (*wA*) is involved in the process of conidial development in *A. niger* and is required for pigment synthesis. Deletion of *wA* leads to *A. niger* colonies with white or fawn colored spores (Jorgensen et al. 2011b; Zhang et al. 2016). As the *wA* gene of LDM3 carries a non-synonymous SNP mutation, it is worth investigating in more detail in the future. In view of the above analysis, we finally focused on two genes, *tupA* and *prpA* to determine whether they mediate both the aconidial and high-secretion phenotypes in *A. niger*, and we selected a conidial lab strain as the recipient for knockout studies. Our data suggested that the knockout of *prpA* barely affected sporulation and protein secretion, while the knockout of *tupA* reduced growth rate and sporulation, that was paralleled by increased secretion capacities. Analysis of publicly available transcriptome data for Δ*tupA* in *A. niger* (Schachtschabel et al. 2013) uncovered that the majority of PrtT-dependent proteases were significantly up-regulated during the early exponential phase, for instance, the expression of *pepA* and *pepB* showed an up to 224- and 99-fold increase, respectively. In addition, several GO terms associated with amino acid biosynthetic or metabolic processes were down-regulated, such as branched-chain family amino acid biosynthetic processes, coenzyme biosynthetic process, and protein secretion (Schachtschabel et al. 2013). This expression data indicates that the low-level protein production during the early fermentation phase of Δ*tupA* compared to the reference strain might be due to the accumulation of extracellular proteases and weakened amino acid biosynthesis. It has been reported that a liquid culture of *A. niger* conidiates abundantly

at the cost of protein secretion when it enters carbon starvation (Jorgensen et al. 2011a). The experimental results shown in Fig. 6 matches this observation, further suggesting that reduced conidiation may increase protein production in *A. niger* in parallel. However, thorough transcriptome and secretome analyses from bioreactor samples are necessary to prove or disprove this assumption.

In conclusion, comparative genome analysis in this study revealed several hundreds of unique and mutated genes in LDM3, some of which might be associated with its aconidial phenotype and thus its high GlaA secretion capacities. These data highlight that novel hypotheses regarding the link between sporulation and protein secretion in *A. niger* can be gleaned from comparative genomics. Other potential leading genes can be leveraged in the future for systematic optimization of protein production capacities in different *A. niger* strains.

Notes

Authors' contributions

YFS, LMO and SC performed the genomics analyses and executed in silico quality analyses. YFS and TS constructed the co-expression networks. YFS and TS generated deletion strains and characterized them. YPZ and VM initiated this study and coordinated the project. YFS, LMO and VM co-wrote the final text. All authors read and approved the final manuscript.

Ethics approval and consent to participate

This article does not contain any studies with human participants or animals performed by any of the authors.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing financial interests.

References

Adams TH, Bolan MT, Timberlake WE (1988) BrlA is necessary and sufficient to direct conidiophore development in *Aspergillus nidulans*. Cell 54**:** 353-362. doi: 10.1016/0092-8674(88)90198-5

Andersen MR, Salazar MP, Schaap PJ, van de Vondervoort PJ, Culley D, Thykaer J, Frisvad JC, Nielsen KF, Albang R, Albermann K, Berka RM, Braus GH, Braus-Stromeyer SA, Corrochano LM, Dai Z, van Dijck PW, Hofmann G, Lasure LL, Magnuson JK, Menke H, Meijer M, Meijer SL, Nielsen JB, Nielsen ML, van Ooyen AJ, Pel HJ, Poulsen L, Samson RA, Stam H, Tsang A, van den Brink JM, Atkins A, Aerts A, Shapiro H, Pangilinan J, Salamov A, Lou Y, Lindquist E, Lucas S, Grimwood J, Grigoriev IV, Kubicek CP, Martinez D, van Peij NN, Roubos JA, Nielsen J, Baker SE (2011) Comparative genomics of citric acid producing *Aspergillus niger* ATCC 1015 versus enzyme-producing CBS 513.88. Genome Res 21**:** 885-897. doi: 10.1101/gr.112169.110

Arentshorst M, Ram AFJ, Meyer V (2012) Using non-homologous end-joining-deficient strains for functional gene analyses in filamentous fungi. Methods Mol Biol. 835**:** 133–150. doi: doi: 10.1007/978-1-61779-501-5_9.

Ashburner M, Ball CA, Blake JA, Botstein D, Cherry JM (2000) Gene ontology: Tool for the unification of biology. Nature Genetics 25**:** 25-29. doi: 10.1038/75556

Aufauvre-Brown A, Mellado E, Gow N, Holden D (1997) *Aspergillus fumigatus chsE*: A gene related to *CHS3* of *Saccharomyces cerevisiae* and important for hyphal growth and conidiophore development but not pathogenicity. Fungal Genet Biol 21**:** 141-152. doi: 10.1006/fgbi.1997.0959

Baker SE (2006) *Aspergillus niger* genomics: past, present and into the future. Med Mycol 44 Suppl 1**:** S17-21. doi: 10.1080/13693780600921037

Basenko EY, Pulman JA, Shanmugasundram A, Harb OS, Crouch K, Starns D, Warrenfeltz S, Aurrecoechea C, Stoeckert CJJ, Kissinger JC, Roos DS, Hertz-Fowler C (2018) FungiDB: An integrated bioinformatic resource for fungi and oomycetes. J Fungi (Basel) 4**:** 39-67. doi: 10.3390/jof4010039

Boecker S, Grätz S, Kerwat D, Adam L, Schirmer D, Richter L, Schütze T, Petras D, Süssmuth RD, Meyer V (2018) *Aspergillus niger* is a superior expression host for the production of bioactive fungal cyclodepsipeptides. Fungal Biol Biotechnol 5**:** 4-17. doi: doi: 10.1186/s40694-018-0048-3. eCollection 2018

Cairns TC, Nai C, Meyer V (2018) How a fungus shapes biotechnology: 100 years of *Aspergillus niger* research. Fungal Biol Biotechnol 5**:** 13-27. doi: 10.1186/s40694-018-0054-5

Cairns TC, Zheng X, Zheng P, Sun J, Meyer V (2019) Moulding the mould: understanding and reprogramming filamentous fungal growth and morphogenesis for next generation cell factories. Biotechnol Biofuels 12**:** 77-96. doi: 10.1186/s13068-019-1400-4

Carvalho ND, Arentshorst M, Kwon MJ, Meyer V, Ram AF (2010) Expanding the *ku70* toolbox for filamentous fungi: establishment of complementation vectors and recipient strains for advanced gene analyses. Appl Microbiol Biotechnol 87**:** 1463-1473. doi: 10.1007/s00253-010-2588-1

Carvalho ND, Jørgensen TR, Arentshorst M, Nitsche BM, van den Hondel CA, Archer DB, Ram AF (2012) Genome-wide expression analysis upon constitutive activation of the HacA bZIP transcription factor in *Aspergillus niger* reveals a coordinated cellular response to counteract ER stress. BMC Genomics 13**:** 350-367. doi: 10.1186/1471-2164-13-350

Chen CC, Ghaffari N, Qian X, Yoon BJ (2017) Optimal hybrid sequencing and assembly: Feasibility conditions for accurate genome reconstruction and cost minimization strategy. Comput Biol Chem 69**:** 153-163. doi: 10.1016/j.compbiolchem.2017.03.016

de Vries RP, Riley R, Wiebenga A, Aguilar-Osorio G, Amillis S, Uchima CA, Anderluh G, Asadollahi M, Askin M, Barry K, Battaglia E, Bayram O, Benocci T, Braus-Stromeyer SA, Caldana C, Canovas D, Cerqueira GC, Chen F, Chen W, Choi C, Clum A, Dos Santos RA, Damasio AR, Diallinas G, Emri T, Fekete E, Flipphi M, Freyberg S, Gallo A, Gournas C, Habgood R, Hainaut M, Harispe ML, Henrissat B, Hilden KS, Hope R, Hossain A, Karabika E, Karaffa L, Karanyi Z, Krasevec N, Kuo A, Kusch H, LaButti K, Lagendijk EL, Lapidus A, Levasseur A, Lindquist E, Lipzen A, Logrieco AF, MacCabe A, Makela MR, Malavazi I, Melin P, Meyer V, Mielnichuk N, Miskei M, Molnar AP, Mule G, Ngan CY, Orejas M, Orosz E, Ouedraogo JP, Overkamp KM, Park HS, Perrone G, Piumi F, Punt PJ, Ram AF, Ramon A, Rauscher S, Record E, Riano-Pachon DM, Robert V, Rohrig J, Ruller R, Salamov A, Salih NS, Samson RA, Sandor E, Sanguinetti M,

Schutze T, Sepcic K, Shelest E, Sherlock G, Sophianopoulou V, Squina FM, Sun H, Susca A, Todd RB, Tsang A, Unkles SE, van de Wiele N, van Rossen-Uffink D, Oliveira JV, Vesth TC, Visser J, Yu JH, Zhou M, Andersen MR, Archer DB, Baker SE, Benoit I, Brakhage AA, Braus GH, Fischer R, Frisvad JC, Goldman GH, Houbraken J, Oakley B, Pocsi I, Scazzocchio C, Seiboth B, vanKuyk PA, Wortman J, Dyer PS, Grigoriev IV (2017) Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. Genome Biol 18: 28-72. doi: 10.1186/s13059-017-1151-0

Fiedler MRM, Cairns TC, Koch O, Kubisch C, Meyer V (2018a) Conditional expression of the small GTPase ArfA impacts secretion, morphology, growth, and actin ring position in *Aspergillus niger*. Front Microbiol 9: 878-894. doi: 10.3389/fmicb.2018.00878

Fiedler MRM, Barthel L, Kubisch C, Nai C, Meyer V (2018b) Construction of an improved *Aspergillus niger* platform for enhanced glucoamylase secretion. Microb Cell Fact 17: 95-106. doi: 10.1186/s12934-018-0941-8

Galperin MY, Makarova KS, Wolf YI, Koonin EV (2014) Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Research 43: D261-269.

Gillingham AK, Tong AHY, Boone C, Munro S (2004) The GTPase Arf1p and the ER to Golgi cargo receptor Erv14p cooperate to recruit the golgin Rud3p to the cis-Golgi. J Cell Biol 167: 281-292. doi: 10.1083/jcb.200407088

Gong W, Cheng Z, Zhang H, Liu L, Gao P, Wang L (2016) Draft genome sequence of *Aspergillus niger* strain An76. Genome Announc 4: e01700-01715. doi: 10.1128/genomeA.01700-15

Guillemette T, van Peij N, Goosen T, Lanthaler K, Robson GD, van den Hondel CA, Stam H, Archer DB (2007) Genomic analysis of the secretion stress response in the enzyme-producing cell factory *Aspergillus niger*. BMC Genomics 8: 158. doi: 10.1186/1471-2164-8-158

Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ von Mering C, Bork P (2015) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Research 44: D286-293. doi: 10.1093/nar/gkv1248

Jimenez-Ortigosa C, Aimanianda V, Muszkieta L, Mouyna I, Alsteens D, Pire S, Beau R, Krappmann S, Beauvais A, Dufrene YF, Roncero C, Latge JP (2012) Chitin synthases with a myosin motor-like domain control the resistance of *Aspergillus fumigatus* to echinocandins. Antimicrob Agents Chemother 56: 6121-6231. doi: 10.1128/AAC.00752-12

Jorgensen TR, Goosen T, Hondel CA, Ram AF, Iversen JJ (2009) Transcriptomic comparison of *Aspergillus niger* growing on two different sugars reveals coordinated regulation of the secretory pathway. BMC Genomics 10: 44. doi: 10.1186/1471-2164-10-44

Jorgensen TR, Nielsen KF, Arentshorst M, Park J, van den Hondel CA, Frisvad JC, Ram AF (2011a) Submerged conidiation and product formation by *Aspergillus niger* at low specific growth rates are affected in aerial developmental mutants. Appl Environ Microbiol 77: 5270-5277. doi: 10.1128/AEM.00118-11

Jorgensen TR, Park J, Arentshorst M, van Welzen AM, Lamers G, Vankuyk PA, Damveld RA, van den Hondel CA, Nielsen KF, Frisvad JC, Ram AF (2011b) The molecular and genetic basis of conidial pigmentation in *Aspergillus niger*. Fungal Genet Biol 48: 544-553. doi: 10.1016/j.fgb.2011.01.005

Krijgsheld P, Nitsche BM, Post H, Levin AM, Muller WH, Heck AJ, Ram AF, Altelaar AF, Wosten HA (2013) Deletion of *flbA* results in increased secretome complexity and reduced secretion heterogeneity in colonies of *Aspergillus niger*. J Proteome Res 12: 1808-1819. doi: 10.1021/pr301154w

Kumar P, Satyanarayana T (2009) Microbial glucoamylases: characteristics and applications. Crit Rev Biotechnol 29: 225-255. doi: 10.1080/07388550903136076

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL (2004) Versatile and open software for comparing large genomes. Genome Biol 5: R12. doi: 10.1186/gb-2004-5-2-r12

Kwon MJ, Jørgensen TR, Nitsche BM, Arentshorst M, Park J, Ram AF, Meyer V (2012) The transcriptomic fingerprint of glucoamylase over-expression in *Aspergillus niger*. BMC Genomics 13: 701-718. doi: 10.1186/1471-2164-13-701

Lee BN, Adams TH (1994) Overexpression of *flbA*, an early regulator of *Aspergillus* asexual sporulation, leads to activation of *brlA* and premature initiation of development. Mol Microbiol 14: 323-334. doi: 10.1111/j.1365-2958.1994.tb01293.x

Lee BN, Adams TH (1996) *FluG* and *flbA* function interdependently to initiate conidiophore development in *Aspergillus nidulans* through *brlA* beta activation. EMBO J 15: 299-309. doi: 10.1002/j.1460-2075.1996.tb00360.x

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26: 589-595. doi: 10.1093/bioinformatics/btp698

Liu X, Ashforth E, Ren B, Song F, Dai H, Liu M, Wang J, Xie Q, Zhang L (2010) Bioprospecting microbial natural product libraries from the marine environment for drug discovery. J Antibiot 63: 415-422. doi: 10.1038/ja.2010.56

Liu X, Bolla K, Ashforth EJ, Zhuo Y, Gao H, Huang P, Stanley SA, Hung DT, Zhang L (2012) Systematics-guided bioprospecting for bioactive microbial natural products. Antonie Van Leeuwenhoek 101: 55-66. doi: 10.1007/s10482-011-9671-1

Mah JH, Yu JH (2006) Upstream and downstream regulation of asexual development in *Aspergillus fumigatus*. Eukaryot Cell 5: 1585-1595. doi: 10.1128/EC.00192-06

Mellado E, Aufauvre-Brown A, Gow N, Holden D (1996a) The *Aspergillus fumigatus chsC* and *chsG* genes encode class III chitin synthases with different functions. Mol Microbiol 20: 667–679. doi: 10.1007/BF00400554

Mellado E, Specht CA, Robbins PW, Holden DW (1996b) Cloning and characterization of *chsD*, a chitin synthase-like gene of *Aspergillus fumigatus*. FEMS Microbiol Lett 143: 69-76. doi: 10.1111/j.1574-6968.1996.tb08463.x

Mellado E, Dubreucq G, Mol P, Sarfati J, Paris S, Diaquin M, Holden DW, Rodriguez-Tudela JL, Latge JP (2003) Cell wall biogenesis in a double chitin synthase mutant (*chsG-/chsE-*) of *Aspergillus fumigatus*. Fungal Genet Biol 38: 98-109. doi: 10.1016/s1087-1845(02)00516-9

Minoru K, Yoko S, Masayuki K, Miho F, Mao T (2015) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44: D457-462. doi: 10.1093/nar/gkv1070

Mouyna I, Morelle W, Vai M, Monod M, Lechenne B, Fontaine T, Beauvais A, Sarfati J, Prevost MC, Henry C, Latge JP (2005) Deletion of *GEL2* encoding for a beta(1-3)glucanosyltransferase affects morphogenesis and virulence in *Aspergillus fumigatus*. Mol Microbiol 56: 1675-1688. doi: 10.1111/j.1365-2958.2005.04654.x

Nakamura T, Maeda Y, Tanoue N, Makita T, Kato M, Kobayashi T (2006) Expression profile of amylolytic genes in *Aspergillus nidulans*. Biosci Biotechnol Biochem 70: 2363-2370. doi: 10.1271/bbb.50694

Park J, Park J, Jang S, Kim S, Kong S, Choi J, Ahn K, Kim J, Lee S, Kim S, Park B, Jung K, Kim S, Kang S, Lee YH (2008) FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors. Bioinformatics 24: 1024-1025. doi: 10.1093/bioinformatics/btn058

Paul S, Ludena Y, Villena GK, Yu F, Sherman DH, Gutierrez-Correa M (2017) High-quality draft genome sequence of a biofilm forming lignocellulolytic *Aspergillus niger* strain ATCC 10864. Stand Genomic Sci 12: 37-45. doi: 10.1186/s40793-017-0254-2

Pavezzi FC, Carneiro AA, Bocchini-Martins DA, Alves-Prado HF, Ferreira H, Martins PM, Gomes E, da Silva R (2011) Influence of different substrates on the production of a mutant thermostable glucoamylase in submerged fermentation. Appl Biochem Biotechnol 163: 14-24. doi: 10.1007/s12010-010-8963-7

Pel HJ, de Winde JH, Archer DB, Dyer PS, Hofmann G, Schaap PJ, Turner G, de Vries RP, Albang R, Albermann K, Andersen MR, Bendtsen JD, Benen JA, van den Berg M, Breestraat S, Caddick MX, Contreras R, Cornell M, Coutinho PM, Danchin EG, Debets AJ, Dekker P, van Dijck PW, van Dijk A, Dijkhuizen L, Driessen AJ, d'Enfert C, Geysens S, Goosen C, Groot GS, de Groot PW, Guillemette T, Henrissat B, Herweijer M, van den Hombergh JP, van den Hondel CA, van der Heijden RT, van der Kaaij RM, Klis FM, Kools HJ, Kubicek CP, van Kuyk PA, Lauber J, Lu X, van der Maarel MJ, Meulenberg R, Menke H, Mortimer MA, Nielsen J, Oliver SG, Olsthoorn M, Pal K, van Peij NN, Ram AF, Rinas U, Roubos JA, Sagt CM, Schmoll M, Sun J, Ussery D, Varga J, Vervecken W, van de Vondervoort PJ, Wedler H, Wosten HA, Zeng AP, van Ooyen AJ, Visser J, Stam H (2007) Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. Nat Biotechnol 25: 221-231. doi: 10.1038/nbt1282

Perez-de-Nanclares-Arregi E, Etxebeste O (2014) Photo-convertible tagging for localization and dynamic analyses of low-expression proteins in filamentous fungi. Fungal Genet Biol 70: 33-41. doi: 10.1016/j.fgb.2014.06.006

Rhoads A, Au KF (2015) PacBio sequencing and its applications. Genom Proteom Bioinf 13: 278-289. doi: 10.1016/j.gpb.2015.08.002

Rogg LE, Fortwendel JR, Juvvadi PR, Lilley A, Steinbach WJ (2011) The chitin synthase genes *chsA* and *chsC* are not required for cell wall stress responses in the human pathogen *Aspergillus fumigatus*. Biochem Biophys Res Commun 411: 549-554. doi: 10.1016/j.bbrc.2011.06.180

Schachtschabel D, Arentshorst M, Nitsche BM, Morris S, Nielsen KF, van den Hondel CA, Klis FM, Ram AF (2013) The transcriptional repressor TupA in *Aspergillus niger* is involved in controlling gene expression related to cell wall biosynthesis, development, and nitrogen source availability. PLoS One 8: e78102. doi: 10.1371/journal.pone.0078102

Schape P, Kwon MJ, Baumann B, Gutschmann B, Jung S, Lenz S, Nitsche B, Paege N, Schutze T, Cairns TC, Meyer V (2019) Updating genome annotation for the microbial cell factory *Aspergillus niger* using gene co-expression networks. Nucleic Acids Res 47: 559-569. doi: 10.1093/nar/gky1183

Swift RJ, Wiebe MG, Robson GD, Trinci APJ (1998) Recombinant glucoamylase production by *Aspergillus niger* B1 in chemostat and pH auxostat cultures. Fungal Genet Biol 25: 100–109. doi: 10.1006/fgbi.1998.1089

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 45: D362-D368. doi: 10.1093/nar/gkw937

Takeshita N, Ohta A, Horiuchi H (2002) *csmA*, a gene encoding a class V chitin synthase with a myosin motor- like domain of *Aspergillus nidulans*, is translated as a single polypeptide and regulated in response to osmotic conditions. Biochem Biophys Res Commun 298: 103-109. doi: 10.1016/s0006-291x(02)02418-x

Takeshita N, Ohta A, Horiuchi H (2005) CsmA, a class V chitin synthase with a myosin motor-like domain, is localized through direct interaction with the actin cytoskeleton in *Aspergillus nidulans*. Mol Biol Cell 16: 1961-1970. doi: 10.1091/mbc.e04-09-0761

te Biesebeke R, van Biezen N, de Vos W, van den Hondel C, Punt PJ (2005) Different control mechanisms regulate glucoamylase and protease gene transcription in *Aspergillus oryzae* in solid-state and submerged fermentation. Appl Microbiol Biotechnol 67: 75-82. doi: 10.1007/s00253-004-1807-z

van Munster JM, Nitsche BM, Akeroyd M, Dijkhuizen L, van der Maarel MJ, Ram AF (2015) Systems approaches to predict the functions of glycoside hydrolases during the life cycle of *Aspergillus niger* using developmental mutants *brlA* and *flbA*. PLoS One 10: e0116269. doi: 10.1371/journal.pone.0116269

Vesth TC, Nybo JL, Theobald S, Frisvad JC, Larsen TO, Nielsen KF, Hoof JB, Brandl J, Salamov A, Riley R, Gladden JM, Phatale P, Nielsen MT, Lyhne EK, Kogle ME, Strasser K, McDonnell E, Barry K, Clum A, Chen C, LaButti K, Haridas S, Nolan M, Sandor L, Kuo A, Lipzen A, Hainaut M, Drula E, Tsang A, Magnuson JK, Henrissat B, Wiebenga A, Simmons BA, Makela MR, de Vries RP, Grigoriev IV, Mortensen UH, Baker SE, Andersen MR (2018) Investigation of inter- and intraspecies variation through genome sequencing of *Aspergillus* section *Nigri*. Nat Genet 50: 1688-1695. doi: 10.1038/s41588-018-0246-1

Vongsangnak W, Hansen K, Nielsen J (2011) Integrated analysis of the global transcriptional response to alpha-amylase over-production by *Aspergillus oryzae*. Biotechnol Bioeng 108: 1130-1139. doi: 10.1002/bit.23033

Wanka F, Cairns T, Boecker S, Berens C, Happel A, Zheng X, Sun J, Krappmann S, Meyer V (2016) Tet-on, or Tet-off, that is the question: Advanced conditional gene expression in *Aspergillus*. Fungal Genet Biol 89: 72-83. doi: 10.1016/j.fgb.2015.11.003

Wanke C, Eckert S, Albrecht G, van Hartingsveldt W, Punt PJ, van den Hondel CAMJJ, Braus GH (1997) The *Aspergillus niger GCN4* homologue, *cpcA*, is transcriptionally regulated and encodes an unusual leucine zipper. Mol Microbiol 23: 23-33. doi: 10.1046/j.1365-2958.1997.1741549.x

Wieser J, Lee BN, Fondon J, 3rd, Adams TH (1994) Genetic requirements for initiating asexual development in *Aspergillus nidulans*. Curr Genet 27: 62-69. doi: 10.1007/bf00326580

Wolf J, Nicks M, Deitz S, Tuinen Ev, Franzusoff A (1998) An N-end rule destabilization mutant reveals pre-Golgi requirements for Sec7p in yeast membrane traffic. Biochem Biophs Res Commun 243: 191-198. doi: 10.1006/bbrc.1998.8084

Yamada O, Lee BR, Gomi K, Iimura Y (1999) Cloning and functional analysis of the *Aspergillus oryzae* conidiation regulator gene *brZA* by its disruption and misscheduled expression. J Biosci Bioeng 87**:** 424-429. doi: 10.1016/S1389-1723(99)80089-9

Yamashiroy CT, Ebbole DJ, Lee BU, Brown RE, Bourland C, Madi L, Yanofsky C (1996) Characterization of *rco-1* of *Neurospora crassa*, a pleiotropic gene affecting growth and development that encodes a homolog of Tup1 of *Saccharomyces cerevisiae*. Mol Cell Biol 16**:** 6218–6228. doi: 10.1128/mcb.16.11.6218

Yin C, Wang B, He P, Lin Y, Pan L (2014) Genomic analysis of the aconidial and high-performance protein producer, industrially relevant *Aspergillus niger* SH2 strain. Gene 541**:** 107-114. doi: 10.1016/j.gene.2014.03.011

Yin X, Shin HD, Li J, Du G, Liu L, Chen J (2017) Comparative genomics and transcriptome analysis of *Aspergillus niger* and metabolic engineering for citrate production. Sci Rep 7**:** 41040-41057. doi: 10.1038/srep41040

Yuan XL, van der Kaaij RM, van den Hondel CA, Punt PJ, van der Maarel MJ, Dijkhuizen L, Ram AF (2008) *Aspergillus niger* genome-wide analysis reveals a large number of novel alpha-glucan acting enzymes with unexpected expression profiles. Mol Genet Genomics 279**:** 545-561. doi: 10.1007/s00438-008-0332-7

Zhang C, Meng X, Wei X, Lu L (2016) Highly efficient CRISPR mutagenesis by microhomology-mediated end joining in *Aspergillus fumigatus*. Fungal Genet Biol 86**:** 47-57. doi: 10.1016/j.fgb.2015.12.007

Figures

Fig. 1 Synteny map of the scaffolds of *A. niger* LDM3 to the supercontigs of *A. niger* CBS 513.88. The coloring of the scaffolds shows syntenic regions in *A. niger* CBS 513.88. Arabic numerals show the number of the supercontig in *A. niger* CBS 513.88. Grey areas show regions not found in the CBS 513.88 genome sequence. The black line underneath a section of the scaffolds indicates inversion sequence. The blue rectangular shadow across the scaffolds indicates the transposition between the two sequence fragments separated by a red line in the shadow
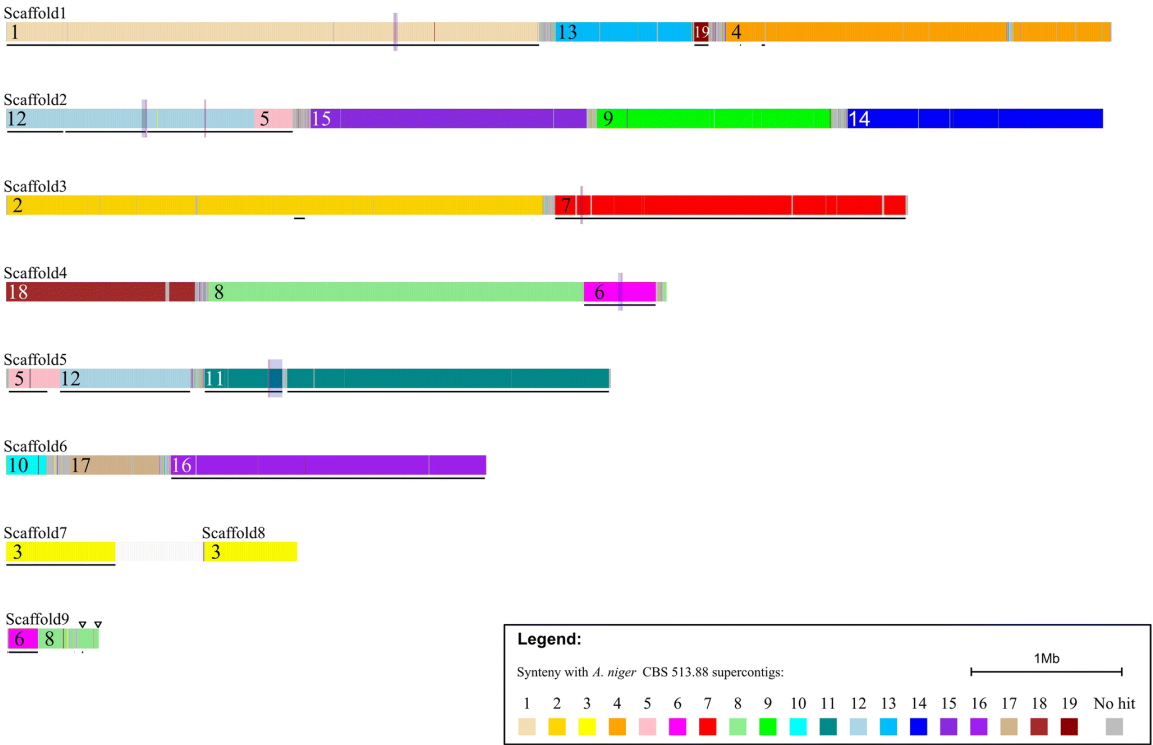
Fig. 2 Distribution of mutated genes in all KOG terms. The height of the histogram shows the total number of genes of *A. niger* LDM3 in various KOG clusters. The grid part represents the number of mutated genes in each KOG category, and the white part means the number of unmutated genes
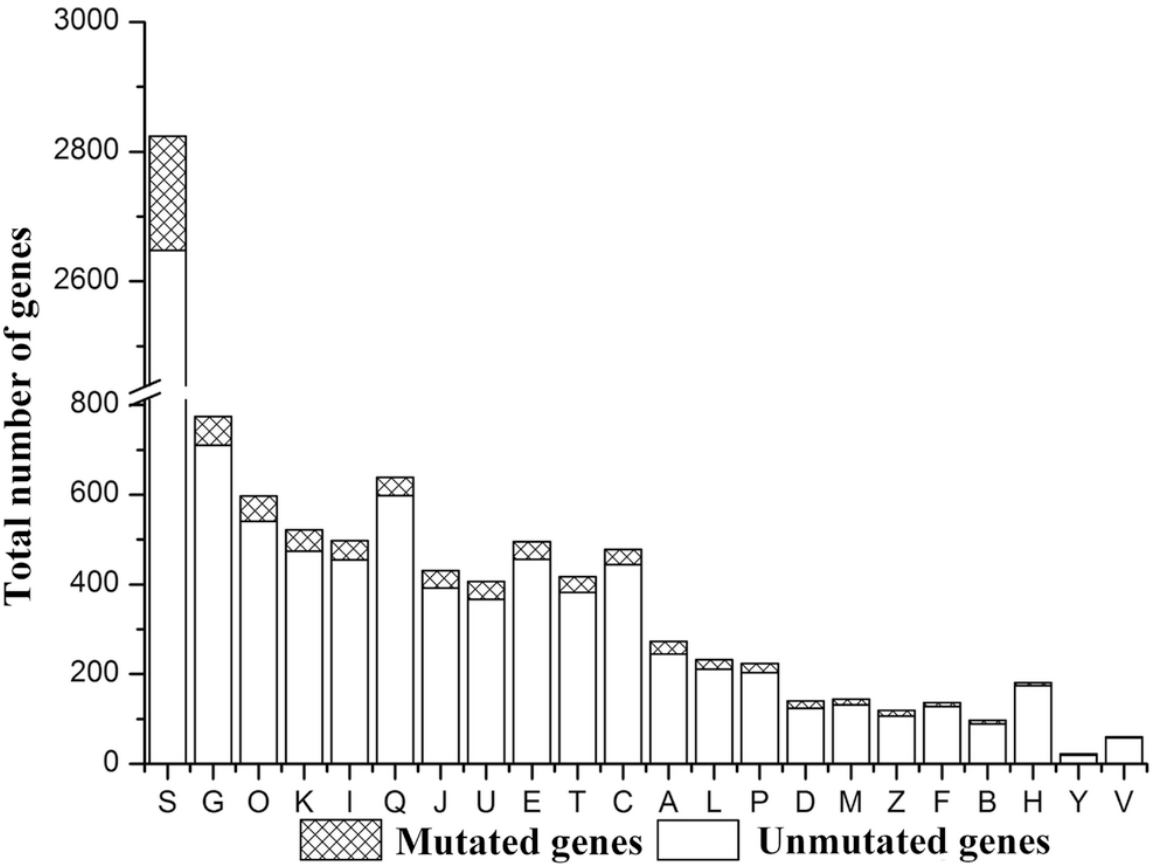
Fig. 3 Co-expression network for TupA and PrpA encoding genes. Protein names are represented by circles, and the query proteins are given in diamond boxes. Only negative correlations for both TupA (a) and PrpA (b) are shown as both of them only express significant negative association with other genes. If gene names were not available in *A. niger*, the name for either the *A. nidulans* or *S. cerevisiae* ortholog was used. Gene pairs within the co-expression sub-networks all passed the |0.7| Spearman correlation coefficient cut-off. Overrepresented GO terms with functions in biological processes above a Benjamini–Hochberg false discovery rate corrected *p* values of 0.05 were assessed, and genes of interest manually filtered
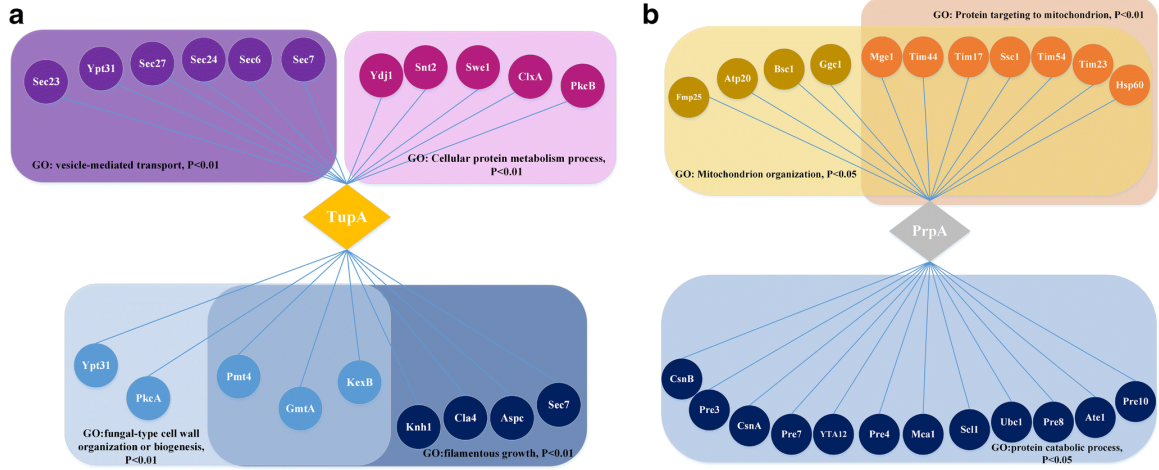
Fig. 4 Phenotypic analysis of Δ*tupA*, Δ*prpA* mutants, and Δ*tupA* complemented strain on solid agar plates. Spores (5*10^4) were spot-inoculated on the different types of media and incubated at 30°C for 3 days. Strain FW35.1 was used as the reference strain. To fully repress any growth, 150 μg/ml hygromycin was adopted. hyg hygromycin
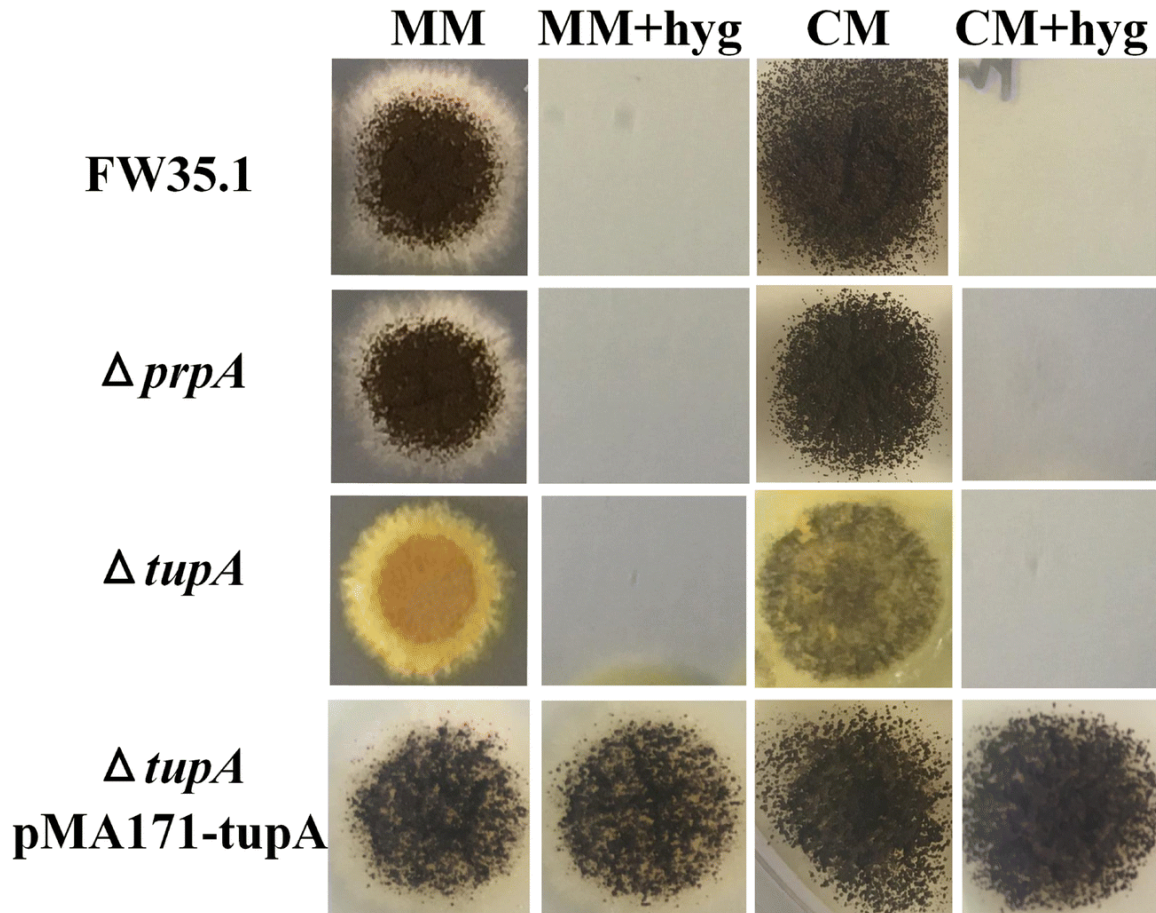
Fig. 5 Sporulation quantification of Δ*tupA*, Δ*prpA* mutants and Δ*tupA* complemented strain on solid agar plates. (a) 1000 spores of each strain were inoculated on the different types of media, respectively, and incubated at 30° C for 6 days. (b) The colony size of different strains after 6 days of culture on CM (stripe) and MM (spot) medium, respectively. (c) The number of spores per square centimeter of interested strains. All experiments were conducted in biological triplicates. Significance values were calculated with the 2-tailed *t*-test with independent variables (*$p<0.05$, **$p<0.01$, ***$p<0.001$)
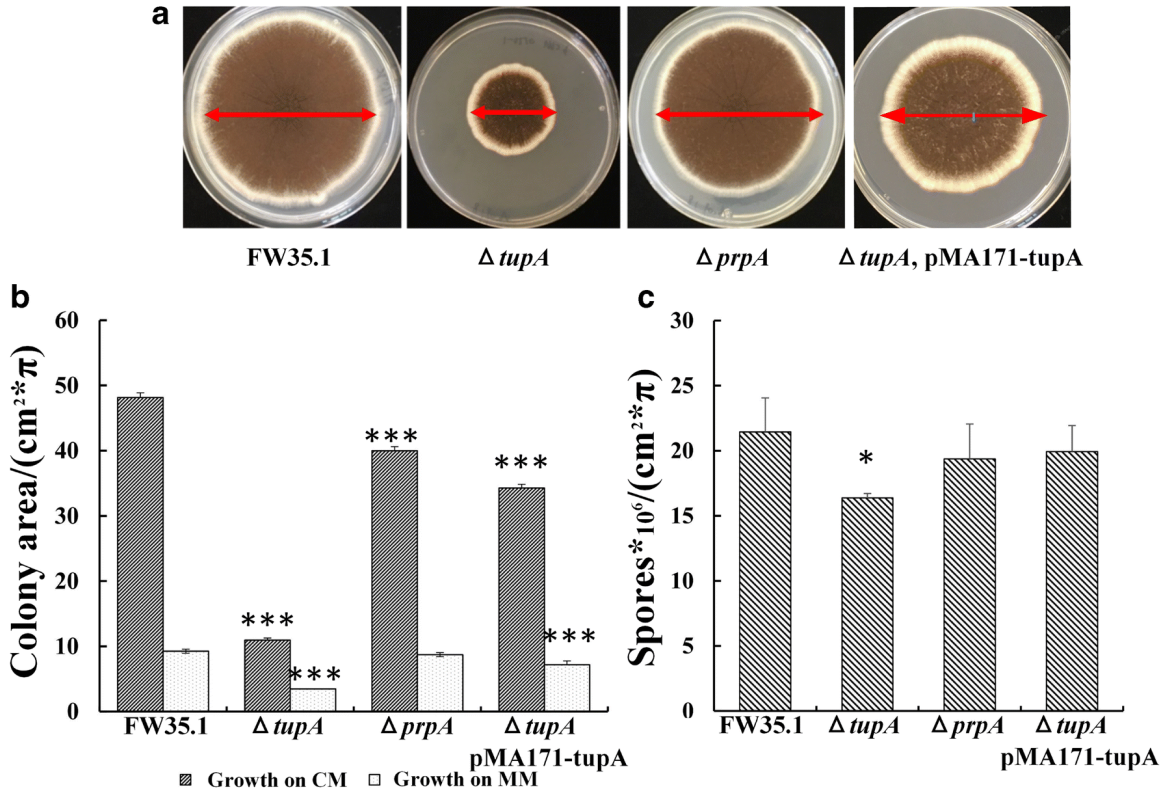
Fig. 6 Characteristics of Δ*tupA* and Δ*prpA* mutants during submerged cultivation. (a) Color change of Δ*tupA* compared to FW35.1 during shake flask cultivation in CM medium. (b) Biomass accumulation in cultures of FW35.1 (square), Δ*prpA* (diamond), and Δ*tupA* strain (triangle). (c) Accumulation of total secreted protein (normalized by per gram biomass) in FW35.1 (black), Δ*prpA* (grey), and Δ*tupA* (spot). (d) Enzyme activity of GlaA (normalized by per gram biomass) in FW35.1 (black), Δ*prpA* (grey) and Δ*tupA* (spot). All experiments were conducted in biological quadruplicates. Significance values were calculated with 2-tailed *t*-test with independent variables (**p*<0.05, ***p*<0.01, ****p*<0.001)
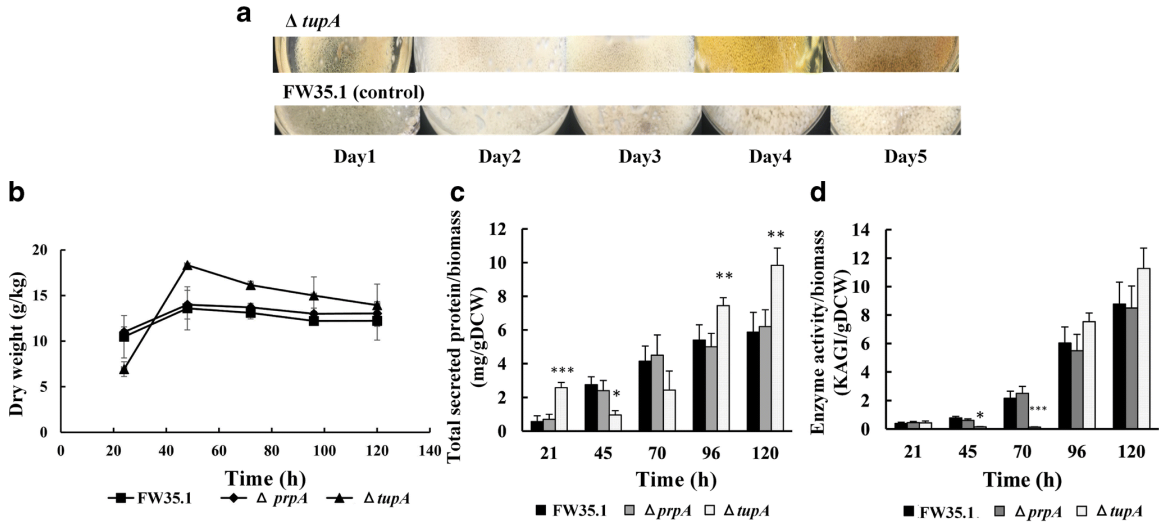
Tables

Table 1 *Aspergillus niger* strains used in this study

Strain name	Background strain	Relevant genotype/description	References
LDM3		Aconidial phenotype	Longda Biotechnology, Shanghai
MA169.4	AB4.1	*cspA1⁻,kusA::DR-amdS-DR, pyrG⁻*	(Carvalho et al. 2010)
FW35.1	AB4.1	*cspA1⁻, pyrG⁺*	(Wanka et al. 2016)
YS33.10	MA169.4	*kusA::DR-amdS-DR, tupA::AopyrG, pyrG⁺*, single copy	This study
YS34.16	MA169.4	*kusA::DR-amdS-DR, prpA::AopyrG pyrG⁺*, single copy	This study
YS39.1	MA169.4	*kusA::DR-amdS-DR, tupA::AopyrG pyrG⁺*, pMA171-*tupA*	This study

Table 2 Genome characteristics of selected *A. niger* strains with industrial relevance

Strain name （NCBI）	ASM285v2	NRRL3	ATCC1015	SH2	LDM3
Name synonym	CBS 513.88	ATCC 9029	NRRL 328		
Accession Number	GCA_000002855.1	unpublished	GCA_000230395.2	GCA_000633045.1	VTFN00000000
Institute and country	DSM, Netherland	Integrated Genomics, USA	DOE/JGI, USA	SCUT, China	ECUST, China
Project Chronology	2000-2007	2000	2005-2011	2013-2014	2016-2017
Genome length	34.02 Mb	33.7 Mb	34.85 Mb	34.63 Mb	35.28 Mb
Sequencing technology	BAC tiling	Shotgun	Shotgun	Illumina HiSeq	Illumina HiSeq + PacBio
Coverage	~7.5 ×	~6×	~8.9 ×	~120 ×	~177 ×
Genomic library insert size	< 150 kb	1-2 kb	3 kb 8 kb 40 kb	500 bp	270 bp (Illumina)
Number of contigs or scaffolds	19 Scaffolds	9510 Contigs	24 Contigs	349 Scaffolds	11 Scaffolds
Number of predicted genes	14,165	14,000	11,200	11,517	11,209

Strain application	GlaA production	Gluconate production	Citric acid production	GlaA production	GlaA production
Reference	(Pel et al. 2007)	(Baker 2006; Vesth et al. 2018)	(Andersen et al. 2011)	(Yin et al. 2014)	This study

Table 3 Selected list of genes present in both LDM3 and CBS 513.88 genomes with SNP or INDEL mutations

Gene ID (CBS 513.88 nomenclature)	Gene name	SNP or InDel	Predicted gene function	Reference
Transcription factors				
An01g07900	*cpcA*	Insertion	Transcription factor important for amino acid biosynthesis under amino acid starvation conditions	(Jorgensen et al. 2009; Pel et al. 2007; Vongsangnak et al. 2011; Wanke et al. 1997; Yin et al. 2014)
An04g06910	*amyR*	Nonsynonymous	Transcription factor for starch hydrolase genes	(Yuan et al. 2008)
An04g06920	*agdA*	Nonsynonymous	Secreted α-glucosidase	
An15g00140	*tupA*	Nonsynonymous	Transcriptional repressor important for cell wall remodelling	(Schachtschabel et al. 2013)
Transporters				
An15g04270		Nonsynonymous	Sugar transporter	
An11g09600		Nonsynonymous	MFS monosaccharide transporter	
An07g03690		Nonsynonymous	Neutral amino acid transporter	
An07g03970		Nonsynonymous	Neutral amino acid transporter	
An14g07130		Nonsynonymous	Neutral amino acid transporter	
Protein secretion and degradation				
An07g02190	*sec7*	Nonsynonymous	Guanyl-nucleotide exchange factor important for intra-Golgi and ER-to-Golgi transport	(Wolf et al. 1998)
An08g00290	*rud3*	Nonsynonymous	Matrix protein of Golgi	(Gillingham et al. 2004)
An07g03880	*pepC*	Nonsynonymous	Subtilisin-like serine protease	
An07g08030	*pepF*	Nonsynonymous	Serine-type carboxypeptidase	
An11g06350	*cpy1*	Nonsynonymous	Serine carboxypeptidases	
Cell wall biosynthesis				
An08g07350	*gelB*	Nonsynonymous	Glucosyltransferase	(Mouyna et al. 2005)
An02g02360	*csmA*	Nonsynonymous	Chitin synthase	(Aufauvre-Brown et al. 1997; Jimenez-Ortigosa et al. 2012; Takeshita et al.

				2002; Takeshita et al. 2005)
Sporulation				
An02g03160	*flbA*	Synonymous	Regulator of G-protein signalling	(Lee and Adams 1994; Perez-de-Nanclares-Arregi and Etxebeste 2014; van Munster et al. 2015; Wieser et al. 1994; Yin et al. 2014)
An12g02050	*wA*	Nonsynonymous	Polyketide synthase important for pigment biosynthesis	(Jorgensen et al. 2011b; Zhang et al. 2016)
An18g01170	*prpA*	Absent	Role in asexual sporulation	(Yin et al. 2014)