# Genome sequencing of the neotype strain CBS 554.65 reveals the MAT1-2 locus of *Aspergillus niger*

**Valeria Ellena**
Austrian Centre of Industrial Biotechnology (Austria)
**Sjoerd J. Seekles**
TiFN
**Arthur F.J. Ram**
Leiden University
**Matthias G. Steiger** ( matthias.steiger@tuwien.ac.at )
TU Wien

**Research Article**

**Keywords:** sexual development, mating-type locus, mitochondrial DNA, centromer

**Posted Date:** March 17th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-256129/v1

## Abstract

# Background

*Aspergillus niger* is a ubiquitous filamentous fungus widely employed as a cell factory thanks to its abilities to produce a wide range of organic acids and enzymes. Due to its economic importance and its role as model organism to study fungal fermentation, its genome was one of the first *Aspergillus* genomes to be sequenced in 2007. Nowadays, the genome sequences of at least five other *A. niger* strains are available. These, however, do not include the neotype strain CBS 554.65.

# Results

In this study, the genome of CBS 554.65 was sequenced with PacBio. A high-quality nuclear genome sequence consisting of 17 contigs with a N50 value of 4.07 Mbp was obtained. The sequencing covered all the 8 centromeric regions of the chromosomes. In addition, a complete circular mitochondrial DNA assembly was obtained. In silico analyses revealed the presence of a MAT1-2-1 gene in this genome, contrary to the so far sequenced *A. niger* strains, which all contain a MAT1-1-1 gene. An alignment at the MAT locus showed a different position of the MAT1-1-1 gene of ATCC 1015 compared to the MAT1-2-1 gene of CBS 554.65, relative to the surrounding genes. In addition, 24 other sequenced isolates of *A. niger* showed a 1:1 ratio of MAT1-1 and MAT1-2 loci. While the genetic organization of the MAT1-2 locus of CBS 554.65 is similar to what is found in other aspergilli, the genetic organization of the MAT1-1 locus is flipped in all sequenced strains.

# Conclusions

This study, besides providing a high-quality genome sequence of an important *A. niger* strain, suggests the occurrence of genetic flipping or switching events at the MAT1-1 locus of *A. niger.* These results provide new insights in the mating system of *A. niger* and could contribute to the investigation and potential discovery of sexuality of this so far asexual fungal species.

## Background

*Aspergillus niger* is a filamentous fungus belonging to the genus *Aspergillus*, section *Nigri*. It exhibits a very versatile metabolism that allows it to grow in different environmental conditions [1]. Since the beginning of the last century *A. niger* has become a major industrial species, used for the production of citric acid, which is employed in many commodity products, including cleaning agents, food and beverages, pharmaceuticals and cosmetics [2, 3]. Moreover, it is used to produce gluconic and fumaric acids [3]. Besides being a good organic acid producer, *A. niger* is also employed for the production of various enzymes, such as amylases, invertase, pectinases, phytases and proteinases, which find applications in the food and feed industry [3]. Being a non-pathogenic and non-toxic organism for healthy individuals, with a long history of safe industrial use, *A. niger* was given the GRAS (generally regarded as safe) status by the United States Food and Drug Administration [3].

In 2007, the genome sequence of the enzyme-producing strain CBS 513.88 was published [4], followed by the sequencing of the citric acid-producing strain ATCC 1015 in 2011 [5]. At the moment, the genome sequences of three other strains, N402, ATCC 13496, NRRL3, are available. Despite its importance as a neotype strain, the *A. niger* strain

CBS 554.65 has not yet been sequenced. This strain was isolated from a tannic-gallic acid fermentation and is denoted as the neotype *A. niger* strain [6]. According to the International Code of Nomenclature for algae, fungi and plants (Shenzhen Code) a neotype is “a specimen or illustration selected to serve as nomenclatural type if no original material exists, or as long as it is missing”[7]. Therefore, the strain CBS 554.65 is used as reference strain for morphological observations and taxonomical studies. A study published in 2016 reported the presence of a MAT1-2-1 gene in the genome of this strain [8], making it an interesting candidate for investigating sexuality in *A. niger*. All the five *A. niger* complete genome sequences reported so far (strains CBS 513.88, ATCC 1015, N402, ATCC 13496 and NRRL3) were shown to contain a MAT1-1-1 gene while the second mating-type locus was not previously described [4, 5]. The MAT loci are regions of the genome [9, 10], which contain one or more open reading frames of which at least one encodes for a transcription factor [10]. Conventionally, the MAT locus containing a transcription factor with an α1 domain similar to the MATα1 of *S. cerevisiae* is called MAT1-1, while the MAT locus containing a transcription factor with a high mobility group (HMG) domain is called MAT1-2 [9]. The corresponding genes are usually called MAT1-1-1 and MAT1-2-1 [9]. The first number indicates that the two sequences are found in the same locus. Due to their sequence dissimilarities they are not termed alleles but, instead, idiomorphs [11]. MAT1-1-1 and MAT1-2-1 are major players in the sexual cycle of fungi. They contain DNA binding motifs and were shown to control the expression of pheromone and pheromone-receptor genes during the mating process [12–14]. In heterothallic species, which are self-incompatible, only one of the two MAT genes is found and mating can occur only between strains of opposite mating-type [9]. In homothallic species, which are self-fertile, both MAT genes are present, either linked or unlinked, in the same genome [15]. In the ascomycetes, the sequences flanking the MAT loci are highly conserved [9, 16, 17]. In the aspergilli, as well as in other fungi, including yeasts, the MAT idiomorphs are usually flanked by the genes *slaB*, encoding for a cytoskeleton assembly control factor, and the DNA lyase *apnB*. Moreover, an anaphase promoting complex gene (*apcE*) is also sometimes present [17].

In this study, we present the full genome sequence of a MAT1-2 *A. niger* strain and compare it to the sequence of 24 *de novo* sequenced *A. niger* isolates containing both MAT1-1 and the MAT1-2 loci.

# Materials And Methods

Strains

The genetic organization of the MAT locus present in *A. niger* CBS 554.65 was analyzed and compared to the MAT locus of *A. niger* ATCC 1015 and 24 *A. niger* isolates obtained from the Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, Utrecht, the Netherlands. The natural isolates analyzed are listed in Table S1 (Additional File 1).

Media

The morphology of strain CBS 554.65 was inspected on minimal medium [18] and malt extract agar (30 g/L malt extract (AppliChem, Darmstadt, Germany) and 5 g/L peptone from casein (Merck KGaA, Darmstadt, Germany)). The strain was 4-point inoculated and incubated at 30°C for one week.

Genome sequencing and annotation

The genome of the *A. niger* neotype strain CBS 554.65 was sequenced with the PacBio® technology by the Vienna Biocenter Core Facilities (VBCF). The genome was assembled with the default HGAP4 pipeline in PacBio SMRTlink. The mitochondrial DNA was assembled using CLC Genomic Workbench 12.0 (QIAGEN). The genome annotation of CBS 554.65 was performed with Augustus [19], first by training the tool on the genome annotation of the strain

ATCC 1015 as reference and then by submitting the genome sequence of CBS 554.65. The genome assembly has been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB42544.

PCRs were performed on the genomic DNA of CBS 554.65 to amplify 1756 bp in the left region (with primers chr5_left_fwd: ACTTATCCCTCGTCAATGA and chr5_left_rev: GGTCGACTTTTTGGGGAAA) and 1638 bp in the right region (with primers chr5_right_fwd_1: TTCTCCATATTGTCAGCCAT and chr5_right_rev_1: CATCGCTTCTTTTCCTCGGA) of chr5_00008F. PCR products were sequenced by Microsynth AG. The MAT locus sequences of 24 *A. niger* isolates were extracted from complete genome sequences obtained with the Illumina technology and assembled using SPADes [20] (data not published). In 18 out of the 24 *A. niger* isolates the MAT locus was distributed over multiple scaffolds. In order to verify the location of the MAT genes and their orientation in these strains, diagnostic PCRs and subsequent sequencing were performed to restore *in silico* gaps within the MAT locus. Primers used for gap restoration are listed in Table S2 (Additional File 2).

*In silico* analyses

The genome sequences of strains ATCC 1015 and NRRL3 were retrieved from JGI [21]. Analyses of the position of the MAT genes within the MAT locus were performed with FungiDB [22] for strains for which a complete genome sequence is available or on BLAST against the whole-genome shotgun contig database (wgs) of *A. niger* for *A. welwitschiae* strains. Sequence analyses and alignments were performed with CLC Main Workbench 8.0.1 (QIAGEN). Homologues of the MAT genes in 24 *A. niger* isolates were determined based on local BlastN searches using genes obtained from CBS 554.65 and ATCC 1015 as query. The sequences of the assembled MAT loci have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB42577.

# Results And Discussion

Morphology of strain CBS 554.65

The strain CBS 554.65 is the *A. niger* neotype, a reference strain for morphological and taxonomical analyses. The morphology of this strain grown on minimal medium and malt extract agar can be observed in Fig. 1. On both media CBS 554.65 forms abundant conidia, black on minimal medium and dark brown on malt extract agar.

Genome sequence and analysis

The genome sequencing of the neotype strain CBS 554.65 yielded 5.3 Gbp in 287,000 subreads. The mean length was 18.4 Kbp for the longest subreads and half of the data was in reads longer than 29 Kbp. The assembly consisted of 17 contigs with a total of 40 Gbp and 55.2-fold coverage. Half of the size of the genome is comprised in 4 scaffolds (L50) of which the smallest has a length of 4.07 Mbp (N50). The GC content is 50.3%. The nuclear genome was annotated with Augustus, using the genome of the strain ATCC 1015 as reference. Based on this automated annotation 12,240 protein coding genes were predicted. In Table 1 some basic characteristics of the CBS 554.65 nuclear genome, calculated with CLC, are reported, in comparison to the characteristics of other three sequenced *A. niger* strains, CBS 513.88, ATCC 1015 and NRRL3, obtained from JGI.

Table 1
Comparison of the basic characteristics of the nuclear genomes of 4 different *A. niger* strains.

| | CBS 554.65 (This study) | CBS 513.88 [4, 5] | ATCC 1015 [5] | NRRL3 [23, 24] |
|---|---|---|---|---|
| Genome size (Mb) | 40.42 | 33.98 | 34.85 | 35.25 |
| Coverage | 55.2x | 7.5x | 8.9x | 10x |
| Number of contigs | 17 | 471 | 24 | 15 |
| Number of scaffolds | 17 | 19 | 24 | 15 |
| Scaffold N50 (Mbp) | 4.07 | 2.53 | 1.94 | 2.81 |
| Scaffold L50 | 4 | 6 | 6 | 5 |
| GC content (%) | 50.3 | 50.4 | 50.3 | 49.92 |
| Protein-coding genes | 12,240 | 14,097 | 11,910 | 11,846 |

The CBS 554.65 genome sequence has an increased quality compared to the sequences of the other strains, with a higher coverage, a higher N50 value and a lower L50 value. Compared to the other sequenced strains, CBS 554.65 appears to have a larger genome, while the GC content is similar in the 4 strains. For each of the 8 chromosomes, a

putative centromeric region between 88 and 100 kb was identified. These are indicated in Figure 2 within two vertical black lines. These regions have a GC content between 17.1 and 18.4%, significantly lower than the GC content characterizing the total genome (50.3%) and do not contain any predicted ORF. The only exception is represented by the centromere of chromosome 1 in which a ORF of 219 nucleotides is annotated. This is found in a 7 kb region of the centromere with higher GC content compared to the GC content of the total centromere, suggesting the presence of a mobile element. A conserved domain search [25] on this sequence gave as hits CHROMO and chromo shadow domains (accession: cd00024), ribonuclease H-like superfamily domain (accession: cl14782), integrase zinc binding domain (accession: pfam17921), reverse transcriptase domain (accession: cd01647), RNase H-like domain found in reverse transcriptase (accession: pfam17919) and a retropepsin-like domain (accession: cd00303). The presence of the last four domains suggests that the analyzed sequence has a retroviral or a retrotransposon origin. Similar sequences with domains for reverse transcriptase were also found in the centromeres of chromosomes 5, 6 and 7. A blast analysis of the single chromosomes of strain CBS 554.65 against the complete genome of strain NRRL3 showed that the putative centromeres are almost completely lacking from the genome assembly of NRRL3 (Figure 2, grey areas in the blast graph). Although difficult to identify, centromeric regions in filamentous fungi are composed of complex and heterogeneous AT rich sequences which can stretch up to 450 kb [26,27]. Due to the likely presence of near-identical long repeats, centromeres are difficult to sequence and assemble [27] explaining why they are lacking in strain NRRL3. Transposon and retrotransposon have been identified in the centromeres of other eukaryotes, including fungi [26,28]. The blast analysis against NRRL3 showed that, besides the putative centromeric regions, other large regions constituting the genome of CBS 554.65 do not find homology in NRRL3, explaining the difference in size between the strains. To confirm that these unique regions are not artifacts, the sequencing reads of CBS 554.65 were remapped to the genome. 192,283 reads were remapped to the genome and the mean read length of the remapped reads was 15,215.97 (see total coverage graph in Figure S1, Additional file 3). High coverage was obtained also for the CBS 554.65 regions which are not found in NRRL3, such as those present in chromosome 4 (chr4_000001F) and chromosome 5 (chr5_000008F) (Figure S2, Additional file 3). Moreover, two PCR reactions could be successfully performed on the non-homologous region in chromosome 5 (indicated by the dots in chr5_000008F, Figure 2). Sequencing of the PCR products confirmed the sequence obtained by genome sequencing. The higher read length obtained with PacBio sequencing allows to cover also repetitive sequences which are probably missing from previous genome sequences of *A. niger* obtained with Illumina, explaining the observed difference in genome size. The number of protein-coding genes in CBS 554.65 is in range with what found in ATCC 1015 and NRRL3. The large difference in the protein-coding genes in strain CBS 513.88 is likely caused by overpredictions, as previously suggested [5]

# Mitochondrial DNA

Many genome projects focused on the nuclear genome while the mitochondrial DNA is often neglected. In *A. niger* only one mitochondrial DNA (mtDNA) assembly has been reported, for the strain N909 [30]. In this study, the mtDNA of strain CBS 554.65 was *de novo* assembled from PacBio reads as a circular DNA with a length of 31,363 bp. MtDNA is abundant in whole genome sequencing projects and the read coverage of the assembly (average: 1,220 x, min: 328 x, max: 1,674 x) is thus higher than for the nuclear genome. In total 18 ORFs, 26 tRNA and 2 rRNA sequences were annotated (Fig. 3). All 15 core mitochondrial genes reported for *Aspergillus* species were identified with a comparable gene organization [31]. In addition, three accessory genes *orf1L*, *orf3* and *endo1* were annotated. The gene *endo1* is located in the intron of *cox1* and encodes a putative homing endonuclease gene belonging to the LAGLIDADG family frequently found in the *cox1* intron of other filamentous fungi [31]. The gene *orf3* encodes for a hypothetical protein of 191 residues, which is also present in the mtDNA of strain N909 but was not annotated there.

Surprisingly this unknown protein has a good hit against an unknown protein of *Staphylococcus aureus* (99% identity), however not against other proteins of *Aspergillus* species. In *A. niger* strain N909 two other unknown proteins are encoded in *orf1* and *orf2*. These two open reading frames are connected to a long one in *A. niger* CBS 554.65 yielding a potential protein product with 739 amino acid residues. This is comparable to an open reading frame located at the same position between *nad1* and *nad4* in the mtDNA of *A. flavus* NRRL 3357 (AFLA_m0040), with a size of 667 amino acid residues. In the N-terminal region of both putative proteins, transmembrane spanning regions can be predicted supposing a location in a mitochondrial membrane, however the C-terminal regions are not conserved between *A. niger* and *A. flavus* protein. It is suggested to use the mitochondrial assembly of CBS 554.65 as a reference sequence for *A. niger* mitochondria because it is known that strain N909 is resistant to oligomycin. This resistance is typically linked to mutations in the mtDNA, either in *atp6* or *atp9*, and indeed two mutations are found in *atp6* of strain N909 (L26W and S173L).

Discovery and sequencing of a MAT1-2 *A. niger* strain

The genome sequencing and analysis of strain CBS 554.65 allowed to determine the mating-type of this strain. The sequence of the putative MAT1-2-1 gene (g9041) was searched in the whole nucleotide database using BlastN, giving as hits the mating-type HMG-box protein MAT1-2-1 of other aspergilli, including *A. neoniger* (with an identity of 93.25%) and *A. tubingensis* (with an identity of 93.07%). As such, we consider gene g9041 to be homologous to the MAT1-2-1 gene of other *Aspergillus* species.

This is in line with a previous study which indicated the presence of a MAT1-2-1 sequence in the CBS 554.65 strain through a PCR approach [8]. Here we report the first complete genome sequence of an *A. niger* strain having a MAT1-2-1 gene. The availability of this genome sequence represents an important tool for further studies investigating the sexual potential of *A. niger*. The presence of both opposite mating-type genes in different strains belonging to the same species represents a strong hint of a sexual lifestyle [10].

MAT1-2 locus analysis and comparison to MAT1-1

The locus of strains CBS 554.65 containing the MAT1-2-1 gene was compared *in silico* to the locus of strain ATCC 1015 containing the MAT1-1-1 gene, to determine whether the genes flanking the MAT1-1-1 gene are also present in the genome of the MAT1-2 strain and vice versa. A region of 40,517 bp, spanning from gene Aspni7|39467 (genomic position 2504615 in the v7 of the ATCC 1015 genome) to gene Aspni7|1128148 (genomic position 2545131) was aligned to the corresponding region of strain CBS 554.65 (Fig. 4). In CBS 554.65 the two genes homologous to Aspni7|39467 (g9051) and Aspni7|1128148 (g9038) are comprised in a sequence of 43,891 bp, almost 4 kb longer than in ATCC 1015. In Fig. 4 genes found in both strains are indicated with a box of the same color, MAT genes are indicated with a circle and genes which are not surrounded by a box or a circle are unique in each strain. The green lines below the genomic region of each strain indicate the sequences homologous in both strains while the black dotted lines indicate the sequences that do not find homology in the other strain. The gene identifiers are indicated on top of each gene and additionally reported in Table 2, with the indication of their predicted function, retrieved from FungiDB or blast analysis. The alignment shows that the MAT genes occupy the same genomic location at chromosome 7. The genes comprised in the analyzed loci are mostly conserved between the two strains, with the exception of genes Aspni7I1178859 (MAT1-1-1), Aspni7I1128137 and Aspni7I1160288, unique for ATCC 1015, and g9046, g9041 (MAT1-2-1) and g9040-2 (MAT1-2-4), unique for CBS 554.65. Aspni7I1128137 has predicted metal ion transport activity and it is found in other *Aspergillus* species, either heterothallic with a MAT1-1-1 or a MAT1-2-1 gene or homothallic, and not in proximity of the MAT gene, with the exception of *A. brasiliensis* and *A. ochraceoroseus*. Aspni7I1160288 has a domain with predicted role in proteolysis and its homolog in other aspergilli

is present at another genomic locus, not in proximity of the MAT gene. A homolog of gene g9046 was found by BlastN search in *Aspergillus vadensis*, in a different location of the genome than the MAT locus. These results suggest that these unique genes are likely not part of the “core” MAT locus. The gene g9040-2 is a putative homolog of the MAT1-2-4 gene in *A. fumigatus*, an additional mating-type gene required for mating and cleistothecia formation [32]. Another difference between ATCC 1015 and CBS 554.65 is represented by the gene putatively encoding for a HAD-like protein. While this gene is complete in CBS 554.65 (g9045), it appears disrupted in ATCC 1015 and, therefore, doubly annotated in this strain (Aspni7|1095364 and Aspni7|1128138). The other genes present in the selected genomic region show a high level of conservation, with a higher syntheny further away from the MAT genes (genes in the purple and blue boxes). Moreover, genes encoding for the DNA lyase a*pnB*, the cytoskeleton control assembly factor *slaB* and the anaphase promoting complex *apcE* are present in both MAT loci. These genes are normally found in the MAT loci of other fungi, including yeast [17], and their presence in the MAT loci of *A. niger* further confirms the high level of conservation characterizing this locus. In heterothallic ascomycetes the MAT genes are commonly included between the genes *apnB* and *slaB* [17]. From the alignment in Fig. 4 the relative position of the MAT genes to *apnB* and *slaB* can be analyzed. In CBS 554.65 the MAT1-2-1 gene (g9041) is flanked by *apnB* and *slaB* respectively upstream and seven genes downstream. In contrast, in the MAT1-1 locus of strain ATCC 1015 the MAT gene is flanked downstream by *apnB* and upstream by a conserved sequence including *adeA*, while *slaB* is found on the same side of *apnB*. The entire genomic locus, containing the MAT1-1-1 gene and eight other genes (23 kbp indicated by the red arrow in Fig. 4), shows a flipped orientation compared to the corresponding locus in CBS 554.65 containing the MAT1-2-1 gene (indicated by an orange arrow in Fig. 4). The ORF direction of the conserved genes *apnB*, *coxM* and *apcE* additionally confirms the different orientation of this locus in the two strains. By sequence analysis, a repetitive 7 bp DNA motif (5´-TTACACT) was found in the MAT1-1 locus (orange triangles in Fig. 4), where the homology between the MAT1-1 and MAT1-2 loci breaks (in proximity to *adeA* and *slaB*). An additional site of this motif was found in the gene encoding a HAD-like hydrolase (Aspni7|1128138). This motif is present at similar positions in two other sequenced MAT1-1 strains of *A. niger* (N402, CBS 513.88). In contrast, the MAT1-2 strain presents this motif only at the site close to the *adeA* gene and in the putative HAD-like hydrolase gene (g9045), but not at the site close to the *slaB* gene.

Table 2
List of genes included in the genomic region comprising the MAT genes.

| ATCC 1015 | CBS 554.65 | Predicted function retrieved from FungiDB or blast |
|---|---|---|
| Aspni7\|39467 | g9051 | Hypothetical protein |
| Aspni7\|1167974 | g9050 | CIA30-domain containing protein – Ortholog(s) have role in mitochondrial respiratory chain complex I assembly |
| Aspni7\|1225150 | g9049 | SAICAR synthetase (*adeA*) |
| Aspni7\|1187920 | g9048 | Homolog in CBS 513.88 has domain(s) with predicted catalytic activity, metal ion binding, phosphoric diester hydrolase activity |
| Aspni7\|39471 | g9040-1 | Hypothetical protein |
| Aspni7\|1178859 | - | Mating-type protein MAT1-1-1 |
| Aspni7\|1187921 | g9042 | DNA lyase Apn2\|Hypothetical protein |
| Aspni7\|1147272 | g9043 | Hypothetical cytochrome C oxidase\|Mitochondrial cytochrome c oxidase subunit VIa |
| Aspni7\|1187923 | g9044 | Ortholog(s) are anaphase-promoting complex proteins |
| Aspni7\|1128137 | - | Homolog in CBS 513.88 has domain(s) with predicted metal ion transmembrane transporter activity, role in metal ion transport, transmembrane transport and membrane localization |
| Aspni7\|1095364 | g9045 | HAD-like protein; Homolog in CBS 513.88 has domain(s) with predicted hydrolase activity |
| Aspni7\|1128138 | g9045 | HAD-like protein; Homolog in CBS 513.88 has domain(s) with predicted hydrolase activity |
| Aspni7\|1187925 | g9047 | Glycosyltransferase Family 8 protein - Ortholog(s) have acetylglucosaminyltransferase activity, role in protein N-linked glycosylation and Golgi medial cisterna localization |
| Aspni7\|1160288 | - | Aspartic protease\|Hypothetical aspartic protease |
| Aspni7\|39480 | g9040 | WD40 repeat-like protein |
| Aspni7\|1187926 | g9039 | Aldehyde dehydrogenase |
| Aspni7\|53077 | g9038 | CoA-transferase family III |
| Aspni7\|1187928 | g9037 | Salicylate hydroxylase |
| Aspni7\|1128148 | g9036 | Cytoskeleton assembly control protein Sla2 |
| - | g9046 | Hypothetical protein |
| - | g9041 | Mating-type HMG-box protein MAT1-2-1 |
| - | g9040-2 | Hypothetical protein – Putative homologue of MAT1-2-4 of *A. fumigatus* |

Methods to identify the opposite mating-type in natural isolates often rely on the use of primers designed to bind to *apnB* and *slaB*, since these are the genes that commonly flank the MAT gene itself [33,34]. In both mating-type *A. niger* strains, *slaB* is found more than 12 kbp far from the MAT gene and this might help explaining why the MAT1-2 locus was never previously described for this species, with only one study mentioning it [8].

Not only the particular orientation of the MAT locus but also the presence of a repetitive motif in the MAT loci suggest that a genetic switch or a flipping event might have occurred or is still ongoing in *A. niger*, which might affect the expression of the MAT genes. Genetic switching events at the MAT locus are known for other ascomycetes, particularly yeasts. For instance, in *S. cerevisiae* a switching mechanism involving an endonuclease and two inactive but intact copies of the MAT genes allows to switch the MAT type of the cell [35]. Expression of the MAT gene is instead regulated in the methylotrophic yeasts *Komagataella phaffii* and *Ogataea polymorpha* via a flip/flop mechanism [36,37]. In these species, a 19 kbp sequence including both mating type genes is flipped so that a MAT gene will be close to the centromere (5 kbp from the centromere) and, therefore, silenced while the other will be transcribed. In CBS 554.65 the region comprising the MAT1-2-1 gene is present at around 280 kbp downstream of the putative centromere, which is much further away of what observed for *K. phaffi* and *O. polymorpha*. However, in certain basidiomycetes, such as *Microbotryum saponariae* and *Microbotryum lagerheimii*, the mating-type locus HD (containing the homeodomain genes) is around 150 kbp distant from the centromere and linked to it [38]. It was proposed that the proximity to the centromere in these species might be enough to reduce recombination events [38]. The effect of the distance between the centromere and the MAT genes in *A. niger* merits further attention, especially in view of a potential sexual cycle happening in this species.

Inversion at the MAT locus have been described for certain homothallic filamentous fungi such as *Sclerotinia sclerotiorum* and *Sclerotinia minor* [39,40]. Field analysis of a large number of isolates showed that strains belonging to these species can either present a non-inverted or an inverted MAT locus. In the inverted orientation two of the four MAT genes at the locus have the opposite orientation and one gene is truncated. In the case of *S. sclerotiorum*, differences in the gene expression were observed between inverted and non-inverted strains. This inversion, induced by crossing-over between two identical inverted repeat present in the locus, likely happens during the sexual cycle before meiosis [39]. The analysis of a larger number of *A. niger* natural isolates is required to investigate whether opposite orientations of both MAT loci exist for this species as well and what the implications of such inversions might be. Chromosomal inversions are considered to prevent recombination between sex determining genes in higher eukaryotes, such as animals and plants [41]. Further studies are therefore required to investigate whether a mechanism similar to those already described in other fungal species is also happening in *A. niger*, which might help to explain the difficulty in finding if this species can bear a sexual cycle.

Genetic comparison of MAT loci in different aspergilli and additional *A. niger* strains

Due to the particular configuration observed in this study for the MAT1-1 locus of strain ATCC 1015, the orientation of the MAT locus of additional *Aspergillus* species for which a genome sequence is available was analyzed (Table 3). Firstly, the genes *adeA* and *slaB* were retrieved because they are conserved and often found at the right and left flank of the MAT gene, respectively (Fig. 4). Subsequently, the position of the MAT gene was checked in comparison to the three conserved genes *apnB*, *coxM* and *apcE*. The MAT gene could be either included between *adeA* and *apnB*, like in ATCC 1015 (flipped position), or between *apnB and slaB*, like in CBS 554.65 (conserved position). The results of this analysis are reported in Table 3. A complete table with the identifiers of all genes analyzed is reported in the Additional file 4.

Table 3
MAT gene identifiers of the analyzed *Aspergillus* strains and their position in the MAT locus.

| Section | Species | Strain | Mating-type gene - MAT | Mating-type | MAT position | Sexual cycle described for the species |
|---|---|---|---|---|---|---|
| *Nigri* | *A. welwitschiae* | CBS 139.54 | 172181 | MAT1-1 | flipped | No |
| | *A. kawachii (A. luchuensis)* | IFO 4308 | AKAW_03832 | MAT1-2 | conserved | No |
| | *A. luchuensis* | 106.47 | ASPFODRAFT_180958 | MAT1-1 | conserved | No |
| | *A. tubingensis* | G131 | Not annotated | MAT1-2 | conserved | Yes [42] |
| | | CBS 134.48 | ASPTUDRAFT_124452 | MAT1-1 | conserved | |
| | *A. niger* | CBS 554.65 | g9041 | MAT1-2 | conserved | No |
| | | ATCC 1015 | ASPNIDRAFT2_1178859 | MAT1-1 | flipped | |
| | *A. brasiliensis* | CBS 101740 | ASPBRDRAFT_167991 | MAT1-2 | flipped | No |
| | *A. carbonarius* | ITEM 5010 | ASPCADRAFT_1991 | MAT1-2 | conserved | No |
| | *A. aculeatus* | ATCC 16872 | ASPACDRAFT_1867751 | MAT1-2 | conserved | No |
| *Nidulantes* | *A. versicolor* | CBS 583.65 | ASPVEDRAFT_82222 | MAT1-2 | conserved | No |
| | *A. sydowii* | CBS 593.65 | ASPSYDRAFT_87884 | MAT1-2 | conserved | No |
| *Ochraceorosei* | *A. ochraceoroseus* | IBT 24754 | P175DRAFT_0477739 | MAT1-1 | conserved | No |
| *Flavi* | *A. flavus* | NRRL 3357 | AFLA_103210 | MAT1-1 | conserved | Yes [43] |
| | *A. oryzae* | BCC7051 | OAory_01101300 | MAT1-2 | conserved | No |
| | | RIB40 | AO090020000089 | MAT1-1 | conserved | |
| *Circumdati* | *A. steynii* | IBT 23096 | P170DRAFT_349471 | MAT1-2 | conserved | No |
| *Candidi* | *A. campestris* | IBT 28561 | P168DRAFT_313902 | MAT1-1 | conserved | No |

| Section | Species | Strain | Mating-type gene - MAT | Mating-type | MAT position | Sexual cycle described for the species |
|---|---|---|---|---|---|---|
| | | | P168DRAFT_285957 | MAT1-2 | conserved | |
| *Terrei* | *A. terreus* | NIH2624 | ATEG_08812 | MAT1-1 | conserved | Yes [44] |
| *Fumigati* | *A. novofumigatus* | IBT 16806 | P174DRAFT_462167 | MAT1-2 | conserved | No |
| | *A. fischeri* | NRRL 181 | NFIA_071100 | MAT1-1 | conserved | Yes [45] |
| | | | NFIA_024390 | MAT1-2 | conserved | |
| | *A. fumigatus* | Af293 | Afu3g06170 | MAT1-2 | conserved | Yes [46] |
| | | A1163 | AFUB_042900 | MAT1-1 | conserved | |
| | | | AFUB_042890 | MAT1-2 | conserved | |
| *Clavati* | *A. clavatus* | NRRL1 | ACLA_034110 | MAT1-1 | conserved | Yes [47] |
| | | | ACLA_034120 | MAT1-2 | conserved | |
| *Aspergillus* | *A. glaucus* | CBS 516.65 | ASPGLDRAFT_89185 | MAT1-1 | n.a.[1] | Yes [48, 49] |
| *Cremei* | *A. wentii* | DTO 134E9 | ASPWEDRAFT_184745 | MAT1-2 | conserved | No |

[1] Conserved genes not in the MAT locus

Table 3. MAT genes included between *adeA* and *apnB* have a flipped orientation while MAT genes included between *apnB* and *slaB* have a conserved orientation. *Aspergillus* species are grouped in sections based on the most updated classification [50]. For each species it is indicated if a sexual cycle was reported.

In the analyzed *Aspergillus* sequences the MAT gene (either MAT1-1-1 or MAT1-2-1) was mostly found between the genes *apnB* and *slaB*, such as in CBS 554.65 (conserved). The only exceptions, showing a configuration similar to the MAT1-1 locus of ATCC 1015, were the MAT1-1-1 gene of *A. welwitschiae* and the MAT1-2-1 gene of *A. brasiliensis*. This analysis could not be performed on the MAT1-2 locus of *A. welwitschiae* nor on the MAT1-1 locus of *A. brasiliensis*, since sequences are not available. Seven of the analyzed species, including the closely related *A. tubingensis*, were reported to bear a sexual cycle. For all of these species, with the exception of *A. glaucus*, for which the conserved genes were not found in proximity of the MAT gene, the conserved position of the MAT gene was observed. These observations suggest that the position of the MAT gene and the orientation of the locus are critical for sexual development to occur.

Since the orientation observed for the MAT1-1 locus of ATCC 1015 might be peculiar for this *A. niger* strain only, additional analyses were performed to determine the orientation of the MAT locus of other 4 available sequenced strains of *A. niger* (CBS 513.88, N402, ATCC 13496, NRRL3) and of natural isolates obtained from various sources. All the *A. niger* strains sequenced so far contain a MAT1-1-1 gene and showed the same orientation of the MAT locus observed in ATTC 1015. In addition, 24 natural isolates of *A. niger* were sequenced and the MAT loci analyzed: 12 contain the MAT1-1 locus and 12 the MAT1-2 locus. The MAT locus configuration of these strains is comparable to the configuration of strain ATCC 1015, in the case of the MAT1-1 strains, and to CBS 554.65, in the case of at least 10 out of 12 MAT1-2 strains. In the two remaining MAT1-2 strains (CBS 118.52 and DTO 175-I5) a gap between two genomic scaffolds could not be closed by PCR, probably because constituting of a region with multiple Gs repeats. However, when aligning the two separate scaffolds of these isolates to the MAT1-2 locus of CBS 554.65, they appear to have the same locus configuration as the other 10 MAT1-2 isolates. Similarly, to what observed for ATCC 1015 and CBS 554.65, the HAD-like protein encoding gene appears disrupted in all the MAT1-1 strains and complete in all the MAT1-2 strains. Further studies are required to investigate whether the disruption of this gene in the MAT1-1 strains plays a role in the context of fungal development. Overall, the MAT 1–1 configuration described in Fig. 4 is a peculiar feature of *A. niger* and its close relative *A. welwitschiae*. Despite showing this unusual orientation, the presence of a 1:1 MAT1-1:MAT1-2 ratio among 24 randomly selected natural *A. niger* isolates is an important observation, which suggests that sexual reproduction is occurring in this species. Moreover, *A. niger* was previously shown to be able to form sclerotia [51–54], important prerequisite for sexual development in closely related species. Therefore, further research should focus on the possibility to efficiently induce a sexual cycle in *A. niger*.

# Conclusions

The annotated genome sequence of CBS 554.65, belonging to the *A. niger* neotype strain, represents an important tool for further studies, considering the high quality of this genome sequence, covering all the 8 centromeres and including a complete mtDNA sequence. The analysis of this genome revealed the presence of a second mating-type locus (MAT1-2) in this strain, making it therefore suitable to investigate fungal development in *A. niger*. The position and the orientation of the MAT1-2-1 gene of *A. niger*, both in the CBS 554.65 strain and in 10 natural isolates, was found to be similar to that of other aspergilli, with the MAT gene included between the genes *apnB* and *slaB*. On the contrary, the unusual position of the MAT1-1-1 gene found in the ATCC 1015 strain and other 12 analyzed natural isolates might indicate that flipping or switching events occurred at the MAT locus. Further research is required to investigate whether this difference in the position of the MAT genes in the opposite mating-type strains could have an effect on the expression of the genes included in this genomic region and, therefore, on the possibility of *A. niger* to reproduce sexually.

# Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and material

The datasets supporting the conclusions of this article are available in the European Nucleotide Archive (ENA) at EMBL-EBI under accession numbers PRJEB42544 [https://www.ebi.ac.uk/ena/browser/view/PRJEB42544] and PRJEB42577 [https://www.ebi.ac.uk/ena/browser/view/PRJEB42577].

### Competing interests' statement

The authors declare that they have no competing interests.


### Funding

The COMET center: acib: Next Generation Bioproduction is funded by BMK, BMDW, SFG, Standortagentur Tirol, Government of Lower Austria und Vienna Business Agency in the framework of COMET - Competence Centers for Excellent Technologies. The COMET-Funding Program is managed by the Austrian Research Promotion Agency FFG.


### Authors' contributions

MGS conceived the study. VE and MGS designed the experiments and the *in silico* analyses. VE performed the experiments and the *in silico* analyses of CBS 554.65 and ATCC 1015. SJS and AFJR provided the sequence data of the MAT loci of the natural isolates. SJS performed the experiments and the *in silico* analyses of the natural isolates. VE prepared the manuscript and all authors provided critical feedback. All authors read and approved the final manuscript.


### Acknowledgements

Not applicable


## References


1. Schuster E, Dunn-Coleman N, Frisvad J, van Dijck P. On the safety of Aspergillus niger - a review. Appl Microbiol Biotechnol 2002;59:426–35. https://doi.org/10.1007/s00253-002-1032-6.
2. Currie JN. The citric acid fermentation of Aspergillus niger. J Biol Chem 1917:15–37.
3. Baker SE, Bennett J. The aspergilli. In: Goldman GH, Osmani SA, editors., CRC Press; 2007. https://doi.org/10.1201/9781420008517.
4. Pel HJ, de Winde JH, Archer DB, Dyer PS, Hofmann G, Schaap PJ, et al. Genome sequencing and analysis of the versatile cell factory Aspergillus niger CBS 513.88. Nat Biotechnol 2007;25:221–31. https://doi.org/10.1038/nbt1282.
5. Andersen MR, Salazar MP, Schaap PJ, van de Vondervoort PJI, Culley D, Thykaer J, et al. Comparative genomics of citric-acid-producing Aspergillus niger ATCC 1015 versus enzyme-producing CBS 513.88. Genome Res 2011;21:885–97. https://doi.org/10.1101/gr.112169.110.
6. Kozakiewicz Z, Frisvad JC, Hawksworth DL, Pitt JI, Samson RA, Stolk AC. Proposal for nomina specifica conservanda and rejicienda in Aspergillus and Penicillium (Fungi). Taxon 1992;41:109. https://doi.org/10.2307/1222500.
7. Turland NJ, Wiersema JH, Barrie FR, Greuter W, Hawksworth DL, Herendeen PS, et al. International Code of Nomenclature for algae, fungi and plants. vol. 159. Koeltz Botanical Books; 2018. https://doi.org/10.12705/Code.2018.

8. Mageswari A, Kim J, Cheon K-H, Kwon S-W, Yamada O, Hong S-B. Analysis of the MAT1-1 and MAT1-2 gene ratio in black koji molds isolated from meju. Mycobiology 2016;44:269–76. https://doi.org/10.5941/MYCO.2016.44.4.269.
9. Debuchy R, Turgeon BG. Mating-type structure, evolution, and function in Euascomycetes. Growth, Differ. Sex. Growth, Di, Berlin/Heidelberg: Springer-Verlag; 2006, p. 293–323. https://doi.org/10.1007/3-540-28135-5_15.
10. Dyer PS, Kück U. Sex and the imperfect fungi. The Fungal Kingdom 2017:193–214. https://doi.org/10.1128/microbiolspec.funk-0043-2017.
11. Metzenberg RL, Glass NL. Mating type and mating strategies in Neurospora. BioEssays 1990;12:53–9. https://doi.org/10.1002/bies.950120202.
12. Lee SC, Ni M, Li W, Shertz C, Heitman J. The evolution of sex: a perspective from the fungal kingdom. Microbiol Mol Biol Rev 2010;74:298–340. https://doi.org/10.1128/MMBR.00005-10.
13. Coppin E, Debuchy R, Arnaise S, Picard M. Mating types and sexual development in filamentous ascomycetes. Microbiol Mol Biol Rev 1997;61:411–28. https://doi.org/10.5424/sjar/2014121-4340.
14. Kück U, Böhm J. Mating type genes and cryptic sexuality as tools for genetically manipulating industrial molds. Appl Microbiol Biotechnol 2013;97:9609–20. https://doi.org/10.1007/s00253-013-5268-0.
15. Pöggeler S. Mating-type genes for classical strain improvements of ascomycetes. Appl Microbiol Biotechnol 2001;56:589–601. https://doi.org/10.1007/s002530100721.
16. Galagan JE, Hynes M, Pain A, Machida M, Purcell S, Peñalva MÁ, et al. Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae. Nature 2005;438:1105–15. https://doi.org/10.1038/nature04341.
17. Dyer PS. Sexual reproduction and significance of MAT in the Aspergilli. In: Heitman J, Kronstad J, Taylor J CL (ed), editor. Sex Fungi. ASM Press, American Society of Microbiology; 2007, p. 123–42. https://doi.org/10.1128/9781555815837.ch7.
18. Arentshorst M, Ram AFJ, Meyer V. Using non-homologous end-joining-deficient strains for functional gene analyses in filamentous fungi. In: Bolton MD, Thomma BPHJ, editors. Plant Fungal Pathog. Methods Protoc., vol. 835. Humana Pre, Totowa, NJ: Humana Press; 2012, p. 133–50. https://doi.org/https://doi.org/10.1007/978-1-61779-501-5_9.
19. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res 2005;33:W465–7. https://doi.org/10.1093/nar/gki458.
20. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19:455–77. https://doi.org/10.1089/cmb.2012.0021.
21. Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, et al. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res 2014;42:D26–31. https://doi.org/10.1093/nar/gkt1069.
22. Basenko E, Pulman J, Shanmugasundram A, Harb O, Crouch K, Starns D, et al. FungiDB: An Integrated Bioinformatic Resource for Fungi and Oomycetes. J Fungi 2018;4:39. https://doi.org/10.3390/jof4010039.
23. Aguilar-Pontes MV, Brandl J, McDonnell E, Strasser K, Nguyen TTM, Riley R, et al. The gold-standard genome of Aspergillus niger NRRL 3 enables a detailed view of the diversity of sugar catabolism in fungi. Stud Mycol 2018;91:61–78. https://doi.org/10.1016/j.simyco.2018.10.001.

24. Vesth TC, Nybo JL, Theobald S, Frisvad JC, Larsen TO, Nielsen KF, et al. Investigation of inter- and intraspecies variation through genome sequencing of Aspergillus section Nigri. Nat Genet 2018;50. https://doi.org/10.1038/s41588-018-0246-1.
25. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res 2020;48:D265–8. https://doi.org/10.1093/nar/gkz991.
26. Smith KM, Galazka JM, Phatale PA, Connolly LR, Freitag M. Centromeres of filamentous fungi. Chromosom Res 2012;20:635–56. https://doi.org/10.1007/s10577-012-9290-3.
27. Friedman S, Freitag M. Centrochromatin of Fungi. Prog Mol Subcell Biol 2017;56:85–109. https://doi.org/10.1007/978-3-319-58592-5_4.
28. Talbert PB, Henikoff S. What makes a centromere? Exp Cell Res 2020;389:111895. https://doi.org/10.1016/j.yexcr.2020.111895.
29. Darling ACE, Mau B, Blattner FR, N.T. P. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:1394–403. https://doi.org/10.1101/gr.2289704.
30. Juhász Á, Pfeiffer I, Keszthelyi A, Kucsera J, Vágvölgyi C, Hamari Z. Comparative analysis of the complete mitochondrial genomes of Aspergillus niger mtDNA type 1a and Aspergillus tubingensis mtDNA type 2b. FEMS Microbiol Lett 2008;281:51–7. https://doi.org/10.1111/j.1574-6968.2008.01077.x.
31. Joardar V, Abrams NF, Hostetler J, Paukstelis PJ, Pakala S, Pakala SB, et al. Sequencing of mitochondrial genomes of nine Aspergillus and Penicillium species identifies mobile introns and accessory genes as main sources of genome size variability. BMC Genomics 2012;13:698. https://doi.org/10.1186/1471-2164-13-698.
32. Yu Y, Amich J, Will C, Eagle CE, Dyer PS, Krappmann S. The novel Aspergillus fumigatus MAT1-2-4 mating-type gene is required for mating and cleistothecia formation. Fungal Genet Biol 2017;108:1–12. https://doi.org/10.1016/j.fgb.2017.09.001.
33. Ramirez-Prado JH, Moore GG, Horn BW, Carbone I. Characterization and population analysis of the mating-type genes in Aspergillus flavus and Aspergillus parasiticus. Fungal Genet Biol 2008;45:1292–9. https://doi.org/10.1016/j.fgb.2008.06.007.
34. Houbraken J, Dyer PS. Induction of the sexual cycle in filamentous ascomycetes. In: van den Berg MA, Maruthachalam K, editors. Genet. Transform. Syst. Fungi, Vol. 2, vol. 2, Cham: Springer International Publishing; 2015, p. 23–46. https://doi.org/10.1007/978-3-319-10503-1_2.
35. Haber JE. Mating-Type Genes and MAT Switching in Saccharomyces cerevisiae. Genetics 2012;191:33–64. https://doi.org/10.1534/genetics.111.134577.
36. Hanson SJ, Byrne KP, Wolfe KH. Flip/flop mating-type switching in the methylotrophic yeast Ogataea polymorpha is regulated by an Efg1-Rme1-Ste12 pathway. PLoS Genet 2017;13:1–26. https://doi.org/10.1371/journal.pgen.1007092.
37. Hanson SJ, Byrne KP, Wolfe KH. Mating-type switching by chromosomal inversion in methylotrophic yeasts suggests an origin for the three-locus Saccharomyces cerevisiae system. Proc Natl Acad Sci U S A 2014;111:E4851–8. https://doi.org/10.1073/pnas.1416014111.
38. Carpentier F, Rodríguez De La Vega RC, Branco S, Snirc A, Coelho MA, Hood ME, et al. Convergent recombination cessation between mating-type genes and centromeres in selfing anther-smut fungi. Genome Res 2019;29:944–53. https://doi.org/10.1101/gr.242578.118.
39. Chitrampalam P, Inderbitzin P, Maruthachalam K, Wu BM, Subbarao K V. The Sclerotinia sclerotiorum Mating Type Locus (MAT) Contains a 3.6-kb Region That Is Inverted in Every Meiotic Generation. PLoS One 2013;8.

https://doi.org/10.1371/journal.pone.0056895.
40. Chitrampalam P, Pryor BM. Characterization of mating type (MAT) alleles differentiated by a natural inversion in Sclerotinia minor. Plant Pathol 2015;64:911–20. https://doi.org/10.1111/ppa.12305.
41. Wright AE, Dean R, Zimmer F, Mank JE. How to make a sex chromosome. Nat Commun 2016;7:12087. https://doi.org/10.1038/ncomms12087.
42. Horn BW, Olarte RA, Peterson SW, Carbone I. Sexual reproduction in Aspergillus tubingensis from section Nigri. Mycologia 2013;105:1153–63. https://doi.org/10.3852/13-101.
43. Horn BW, Moore GG, Carbone I. Sexual reproduction in Aspergillus flavus. Mycologia 2009;101:423–9. https://doi.org/10.3852/09-011.
44. Arabatzis M. Sexual reproduction in the opportunistic human pathogen Aspergillus terreus 2013;105:71–9. https://doi.org/10.3852/11-426.
45. Raper KB, Fennell DI. The genus Aspergillus. 1965.
46. O'Gorman CM, Fuller HT, Dyer PS. Discovery of a sexual cycle in the opportunistic fungal pathogen Aspergillus fumigatus. Nature 2009;457:471–4. https://doi.org/10.1038/nature07528.
47. Ojeda-López M, Chen W, Eagle CE, Gutiérrez G, Jia WL, Swilaiman SS, et al. Evolution of asexual and sexual reproduction in the aspergilli. Stud Mycol 2018;91:37–59. https://doi.org/10.1016/j.simyco.2018.10.002.
48. Link HF. Observationes in ordines plantarum naturales. Dissertatio I. Mag Ges Naturf Freunde Berlin 3 1809;3–42.
49. Chen AJ, Hubka V, Frisvad JC, Visagie CM, Houbraken J, Meijer M, et al. Polyphasic taxonomy of Aspergillus section Aspergillus (formerly Eurotium), and its occurrence in indoor environments and food. Stud Mycol 2017;88:37–135. https://doi.org/10.1016/j.simyco.2017.07.001.
50. Houbraken J, Kocsubé S, Visagie CM, Yilmaz N, Wang XC, Meijer M, et al. Classification of Aspergillus, Penicillium, Talaromyces and related genera (Eurotiales): An overview of families, genera, subgenera, sections, series and species. Stud Mycol 2020;95:5–169. https://doi.org/10.1016/j.simyco.2020.05.002.
51. Jørgensen TR, Burggraaf A-M, Arentshorst M, Schutze T, Lamers G, Niu J, et al. Identification of SclB, a Zn(II)2Cys6 transcription factor involved in sclerotium formation in Aspergillus niger. Fungal Genet Biol 2020;139:103377. https://doi.org/10.1016/j.fgb.2020.103377.
52. Frisvad JC, Petersen LM, Lyhne EK, Larsen TO. Formation of sclerotia and production of indoloterpenes by Aspergillus niger and other species in section Nigri. PLoS One 2014;9:e94857. https://doi.org/10.1371/journal.pone.0094857.
53. Ellena V, Bucchieri D, Arcalis E, Sauer M, Steiger MG. Sclerotia formed by citric acid producing strains of Aspergillus niger: induction and morphological analysis. Fungal Biol 2021. https://doi.org/10.1016/j.funbio.2021.01.008.
54. Jørgensen TR, Nielsen KF, Arentshorst M, Park J, van den Hondel CA, Frisvad JC, et al. Submerged conidiation and product formation by Aspergillus niger at low specific growth rates are affected in aerial developmental mutants. Appl Environ Microbiol 2011;77:5270–7. https://doi.org/10.1128/AEM.00118-11.

# Figures

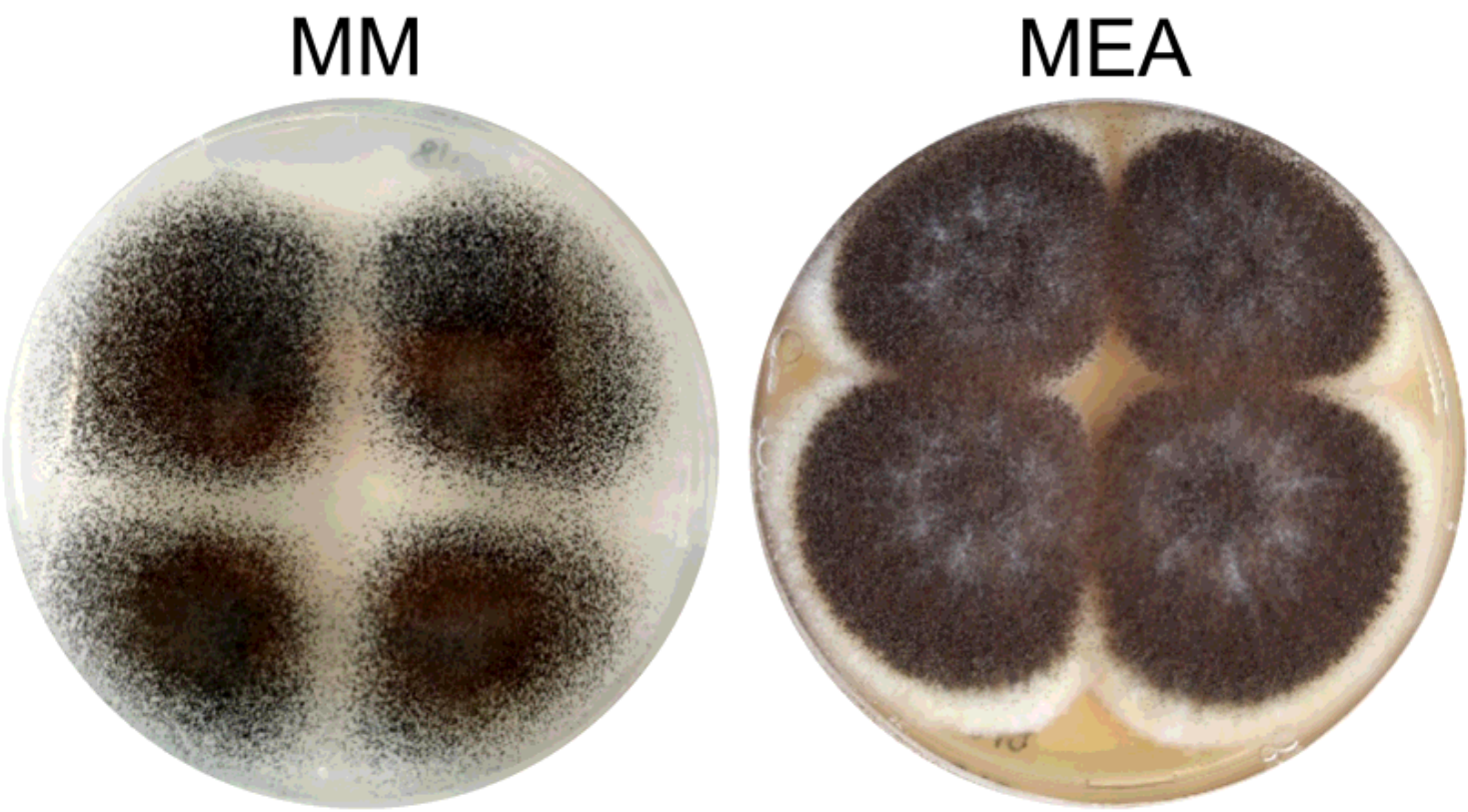


**Figure 1**

Morphology of the neotype strain CBS 554.65 on minimal medium (MM) and malt extract agar (MEA).


## Figure 2

Assembly of the genome sequence of CBS 554.65 consisting of 17 contigs (in scale). For each contig (black horizontal lines) the annotated ORFs (first row), the GC content (second row) and the conservation compared to NRRL3 (third row) are schematically represented. The annotation was obtained with Augustus. The GC content was calculated using a window size of 25 bp. The upper and darker graph represents the maximum GC content value observed in that region, the middle graph represents the mean GC value and the lower graph represents the minimum GC value. The conservation graph (last row) was obtained by blasting each contig of CBS 554.65 against the whole genome of strain NRRL3. The results shown here were additionally confirmed using Mauve [29] by performing progressive alignments of each CBS 554.65 scaffold with the complete genome sequence of NRRL3 (data not shown). Green areas indicate genomic regions conserved between the two strains, grey areas indicate

regions only found in CBS 554.65 and not in NRRL3. Below the conservation graph lines representing the chromosomes of strain NRRL3 are reported, as a result of the blast analysis. Notably, for each of the 8 identified chromosomes, a centromeric region of at least 80 kb could be identified where ORFs are not annotated (indicated with two parallel and vertical lines; the first and the last nucleotide after and before the annotated ORFs, respectively, are indicated). These regions correspond to a decrease in the GC content (as indicated in the GC graph) and are only partially present in the genome of strain NRRL3 (grey areas in the blast graph). Dots on chr5_000008F indicate the region where the PCRs were performed. The MAT locus analyzed in the following paragraphs is indicated by a red box on chromosome 7.

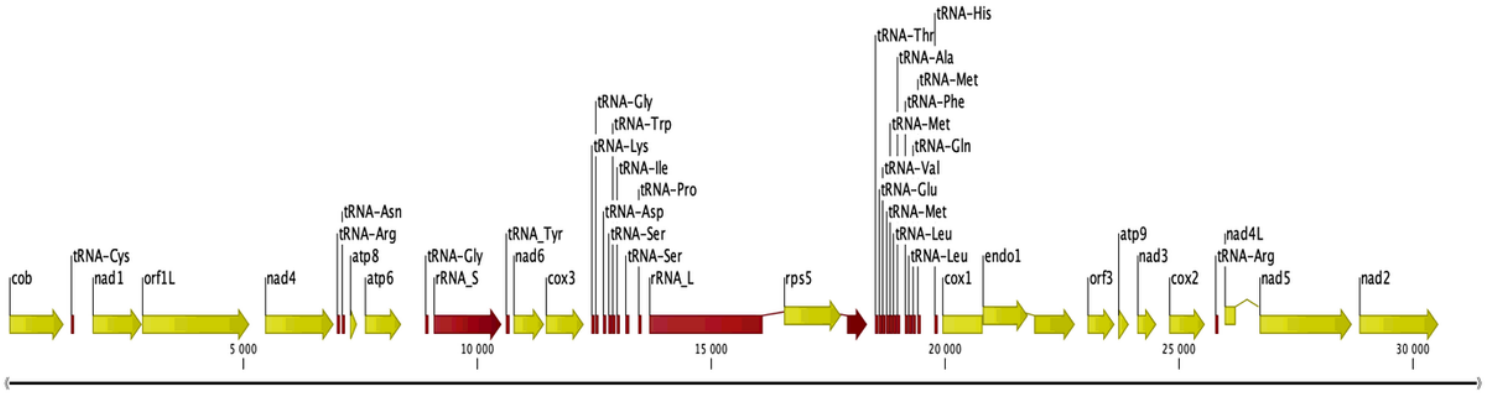


**Figure 3**

Annotation of the 31 kbp circular mtDNA sequence (displayed in a linear projection): ORF (yellow), rRNA, tRNA (red).

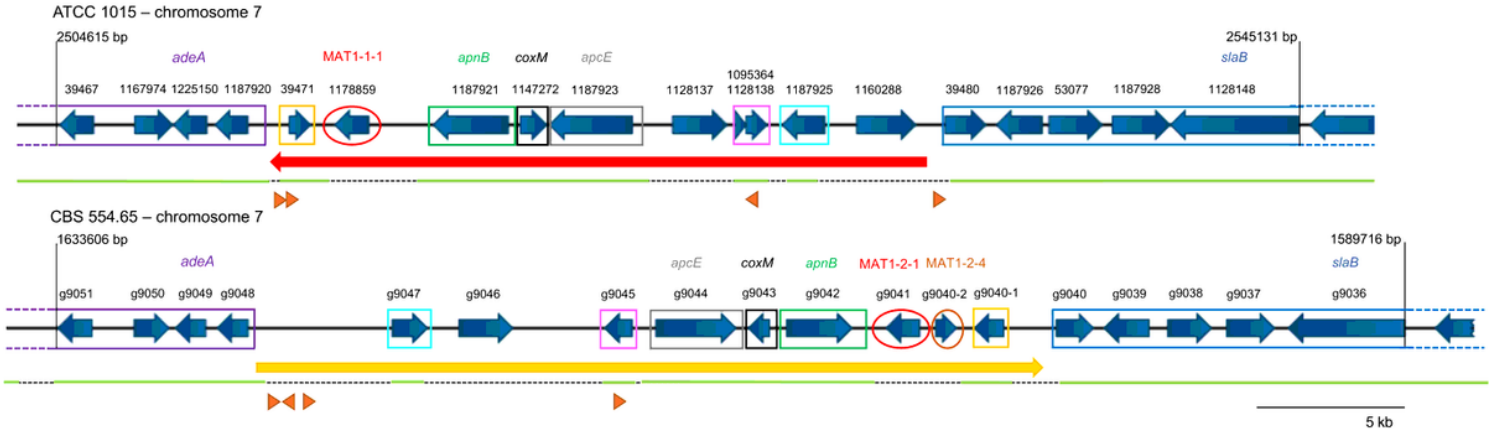


**Figure 4**

Nucleotide alignment between the same genomic region of ATCC 1015 (MAT1-1) and CBS 554.65 (MAT1-2). Boxes of the same color indicate genes found in both strains while circles indicate the MAT genes. Below each genomic region, green lines indicate regions homologous in the two strains and dotted lines regions unique for each strain. A red arrow indicates the genomic region of ATCC 1015 which contains the MAT1-1-1 gene and appears flipped compared to the corresponding region in CBS 554.65 (yellow arrow). Orange triangles indicate the presence of a 7 bp motif (5´-TTACACT).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.pdf
- Additionalfile2.pdf
- Additionalfile3.pdf
- Additionalfile4.pdf