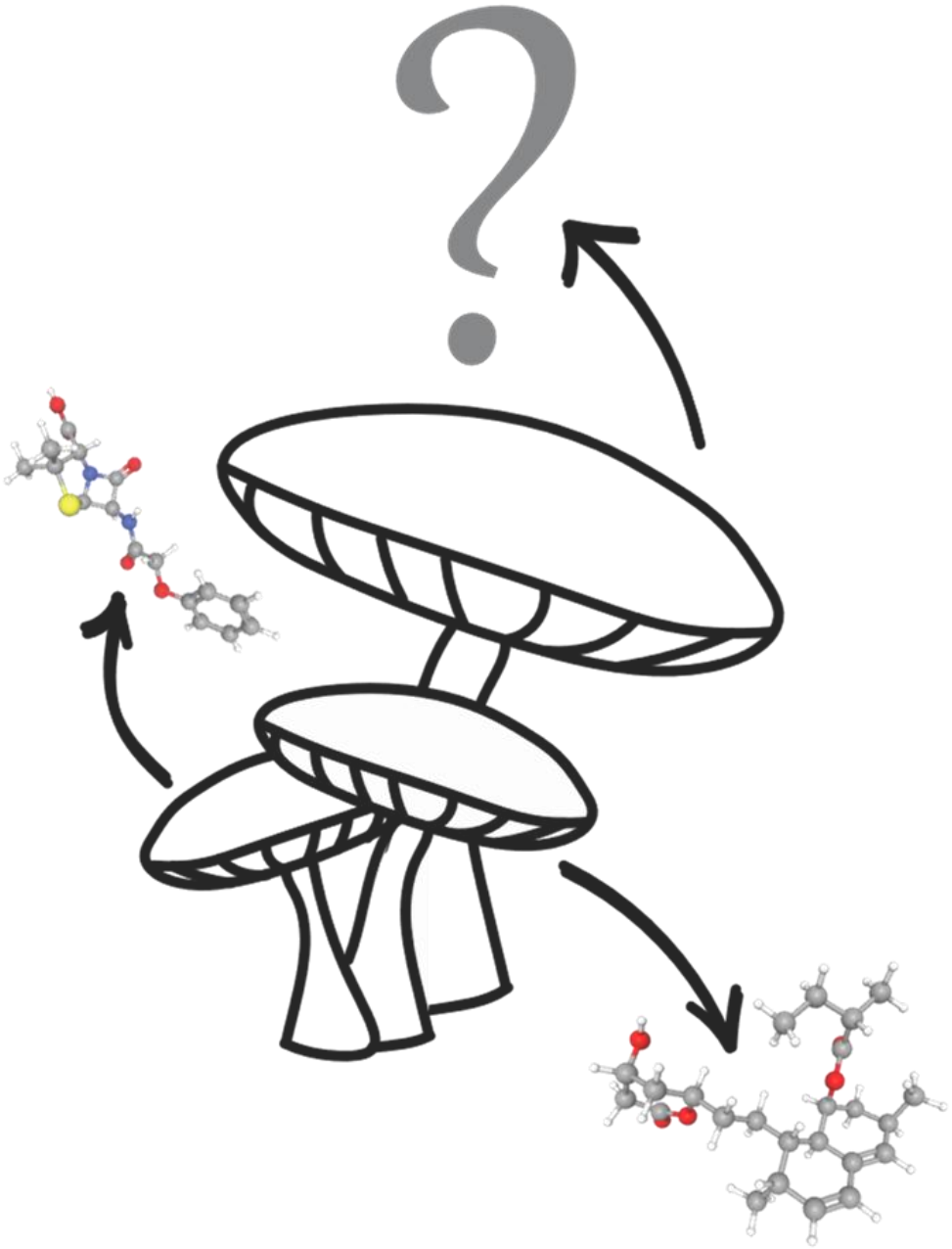
Synthesis of fungal secondary metabolites via heterologous expression



Elske Josefien Dwars 5580978

Daily supervisor: O. (Olga) Mosunova, MSc
Co-supervisor: Dr. J. C. (Jorge) Navarro-Muñoz
First examiner: Dr. Ir. J. (Jérôme) Collemare
Second examiner: Dr. M. F. (Michael) Seidl

Synthesis of fungal secondary metabolites via heterologous expression

Research report of minor internship in the Fungal Natural Products group at the Westerdijk Institute.

Student: Elske Josefien Dwars
Student number: 5580978

Utrecht university – Molecular and cellular life sciences – MSC minor internship
Westerdijk Institute – Fungal Natural Products group

Daily supervisor: O. (Olga) Mosunova, MSc
Co-supervisor: Dr. J. C. (Jorge) Navarro-Muñoz
First examiner: Dr. Ir. J. (Jérôme) Collemare
Second examiner: Dr. M. F. (Michael) Seidl

Table of Contents

Acknowledgements

I would like to extend my sincere gratitude to Olga Mosunova, Jorge C. Navarro-Muñoz, and Jérôme Collemare for their supervision during my internship: your supervision has been extremely helpful. I greatly appreciate your honesty, patience, and your availability whenever I had a question, and I enjoyed the inspiring environment and enthusiasm you shared.

I would also like to thank Michael Seidl for being my examiner from Utrecht University.

Lastly, I would like to thank the whole Fungal Natural Products group. Everyone's kindness and willingness to help makes the group a very nice place to work.

Abstract

In this study, we laid the foundations of a bioinformatic tool to enable for automated gene curation, to facilitate a more correct prediction of fungal gene models. The goal is to employ this tool in the groups of genes (biosynthetic gene clusters, BGCs) coding for enzymes related to the biosynthesis of secondary metabolites (SMs). Correct gene models are crucial for researchers that intend to clone and express BGCs, especially if the pathway of interest is not expressed in a native host under standard laboratory conditions. The gene curation tool can currently split genes to organized blocks representing an exon, and further development will focus on analyses of these blocks to enable automated curation. Additionally, genes from a BGC originating from *Aspergillus melleus* (AspMel) and genes from a BGC originating from *Beauveria bassiana* (BeaBas) were expressed in the heterologous host *A. oryzae* NSAR1. HPLC and LC-MS analysis of the produced and extracted metabolites revealed that the compounds produced by transformants with the AspMel non-reducing polyketide synthase (*AME1*) alone or with *AME1* and tailoring gene *enoyl reductase* absorb light at 213/214 and 261/260 nm respectively, and compounds of the latter, most likely a previously unidentified compound, have a mass of 168 Da. Additionally, the compounds produced by transformants with *AME1* and three tailoring genes (*FAD binding protein, O-methyltransferase*, and *questin oxidase-like*) consistently inhibit growth of *Escherichia coli* and of *Bacillus subtilis*. The transformants with genes from BeaBas did not show any chemical properties or consistent bioactivity.

Laymen´s summary

During their life cycle, fungi produce small molecules called secondary metabolites (SMs). Although called secondary, these metabolites are very important in the survival of fungi, and they mediate interaction with their environment. Often these SMs present biological activities that are valuable for human society, such as antibiotic penicillin, immunosuppressive agent cyclosporin and cholesterol lowering lovastatin. While discovery of bioactive molecules was largely focused on bacteria in the second half of 20th century, recent genomic studies indicate that fungi possess many more SMs, which are not known to science and may be valuable for society. Given the increasing rate of bacterial strains that are resistant to all existing antibiotics, exploration of fungal SMs is essential to provide molecules that can be further developed into medical drugs.

Specific enzymes are responsible for the production of SMs: one or two core enzymes create the primary structure of the SM and then one or more tailoring enzymes modify this primary structure to obtain the final SM. Core enzymes often are very large structures, composed of several well-conserved subsequences, called domains, which each have a distinct function in the synthesis of SMs. The genes coding for core and tailoring enzymes are often co-localized and co-regulated in the fungal genome. Such clusters are called biosynthetic gene clusters (BGC). For a given fungus that contains multiple BGCs, only a fraction of them are active and produce SMs. Some of these BGCs can be activated by changing cultivation conditions, and some remain silent. Additionally, it is difficult to study BGCs from fungi that are difficult to grow in the lab. One way to get to know what SM is assembled by a certain BGC is to transfer (clone) core and tailoring genes into another fungus and make it to assemble the SM of interest; this strategy is called heterologous expression.

Successful cloning and heterologous expression require that all genes that are cloned from one fungus to another, are correctly predicted. Although core and tailoring genes contain well conserved domains, gene models deposited in public databases are often inaccurate, or predicted wrong by automated bioinformatic tools. This especially concerns the start position, intron-exon boundaries, and the end position of a gene. A manual check and manual curation of the gene models is therefore needed to obtain the final, correct genes of the BGC to be able to clone them in a heterologous host. Manual curation of genes is labor intense, requires specific knowledge, and is not feasible for large datasets, but gene model curation is absolutely necessary when preparing genes for cloning.

In this project, two research lines were executed in parallel. One goal regards development of a standalone bioinformatic tool for automatic gene curation. The second goal involves heterologous expression of two unexplored BGCs. Obtained results show that SMs produced by one of these BGCs demonstrate antibiotic activity against gram-positive and gram-negative bacteria.

Introduction

In fungal genomes, primary metabolic pathways are involved in growth, development, and reproduction of fungi. Pathways that are involved in secondary metabolism are associated with the production of small, often bioactive molecules, called secondary metabolites (SMs). In fungal genomes, genes coding for proteins for these pathways are mostly co-localized and co-regulated in the fungal genome [1], [2]. SMs are of great importance for both fungi and human society and exploitation of them could yield novel bioactive compounds, for example in the defense against antibiotic-resistant bacterial strains [3].

1. Fungal secondary metabolites and their impact on human society

SMs exert a wide variety of cellular functions. For example, filamentous fungi produce a wide range of different pigments such as melanin, which are polymers that absorb a broad range of the electromagnetic spectrum. This prevents photoinduced damage caused by ultraviolet light [4]. Also in human society, fungal pigments are widely used, for example as natural colorants in food or textiles [5], [6].

Additionally, several SMs are associated with toxicity, which can also affect the human world. Mycotoxins, for instance, are secondary metabolites that are toxic to vertebrates even in low concentrations [7]. During the Middle Ages, epidemic ergot alkaloid poisoning or ergotism was caused every now and then by consumption of rye bread containing grains infected by the fungus *Claviceps purpurea* [8], [9]. Another group of mycotoxins that are a threat to food safety for both human food and animal feeds, is the aflatoxins [10]. Aflatoxins are produced by *Aspergillus* species and were discovered after the outbreak of a therefor unknown disease of turkeys in England in 1960, called the Turkey "X" Disease. After the outbreak, it was discovered that aflatoxins produced by *Aspergillus flavus*, which had infected the groundnuts used in the turkey feed, had caused this disease [11].

Although mycotoxins can affect human society negatively, other fungal SMs are beneficial. In traditional Chinese medicine, fungi have been used for millennia. Clinical studies have supported the therapeutic value of many fungal products used by traditional Chinese medicine to treat several diseases, such as cancer, diabetes, and neurodegenerative diseases [5]. Secondary metabolites have also revolutionized the modern Western medicine since the discovery of the first large-spectrum antibiotic penicillin, produced by *Penicillium rubens,* in 1928 [12], [13]. The antimicrobials based on SMs discovered and developed since have provided protection against microbial infections for a long time. However, microorganisms develop resistance against such antimicrobials [14]. Therefore, it is important that novel chemical structures are discovered to protect mankind and animals from microbial infections, but since the 1990s, the rate of discovery of such novel bioactive chemical structures has declined. This so-called discovery void was due to the redundant discovery of molecules of natural origin which rendered these of less interest for economical profits. As a result, many pharmaceutical firms eliminated their research for natural products and focused on chemical synthesis of analogs of existing compounds. However, new techniques, including heterologous expression, could allow access to 99% of unexamined microorganisms where novel, undiscovered natural products could be uncovered and may provide a source for many novel antimicrobial compounds in the future [15].

Besides the advantages of fungal SMs in the medicinal world, fungi can also be used in agriculture as biological control agents against pests. Arthropod pests, for example, are a major threat to crop yields which leads to a 15% loss of crops worldwide [16]. Several chemical pesticides are used nowadays to control the pests, but these chemical pesticides can also affect off-target organisms, pollute the environment, and sometimes be found as residues in the final food product. Biopesticides can be used as an alternative to chemical pesticides to reduce the loss of crop yields caused by insect invasions [17], and consist of roughly three categories: plant-incorporated protectants, biochemicals, and microbial pesticides. Entomopathogenic fungi, which fall in the last category, are used to reduce or eradicate specific insect populations that cause a pest [5], [17]. The entomopathogens that show production of SMs upon infection of the host cause rapid host death [18]. This rapid death indicates that the SM might be toxic to target arthropods and could therefore be a potential source of effective and species-specific biochemical pesticides.

2. Expression of fungal biosynthetic gene clusters: source of novel secondary metabolites

A metabolic pathway consists of steps that are necessary for the production of metabolites. Pathway steps reflect chemical rearrangements that can be spontaneous, but mostly are mediated by enzymes. Genes that encode for proteins that synthesize or modify SMs are co-located and co-regulated in fungi in clusters called biosynthetic gene clusters (BGCs); these clusters are a valuable source of novel SMs that may be impactful for human society.

2.1. Biosynthetic gene clusters and their different classes of secondary metabolites

BGCs typically contain one or more core genes, which code for the core enzyme. This core enzyme is responsible for the production of the first stable intermediate, the backbone of the SM. This backbone is then modified by the tailoring enzymes, for example by methylation or decarboxylation [19], [20]. The genes coding for these tailoring enzymes are co-located and co-regulated with the core gene in the BGC (Fig. 1) [21].

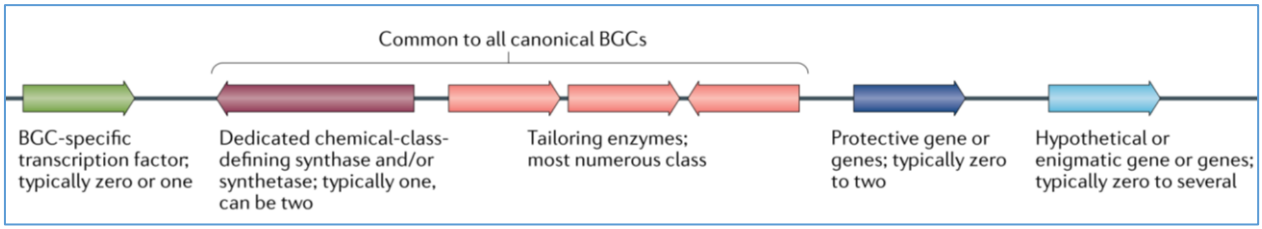


Figure 1. Schematic overview of a biosynthetic gene cluster. Typically, biosynthetic gene clusters (BGCs) contain one or more core genes, coding for the dedicated chemical-class-defining synthase and/or synthase that form the backbone of the secondary metabolite (SM), which is then further modified by the tailoring enzymes, coded for by tailoring genes. Some BGCs also contain a BGC-specific transcription factor, which regulates the other genes within the BGC. Sometimes, protective genes that mitigate the possible toxicity of the SM, or other genes that are not obviously involved in production of or protection from the metabolite can also be present in the BGC. Figure adapted from [12].

The majority of fungal SMs are grouped in five major classes, depending on the core enzymes and on the precursors of the backbone [12], [22] (Fig. 2). These classes include the polyketides, non-ribosomal peptides, terpenes, and indole alkaloids, which are synthesized by polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs), terpene cyclases (TCs), and dimethylallyl tryptophan synthases (DMATSs), respectively [19], [23]. The fifth class includes

hybrid molecules that are produced by hybrid enzymes, such as PKS-NRPS [24], [25]. These core enzymes can be very large structures containing several conserved subsequences, or domains, of which each exerts a specific enzymatic activity.

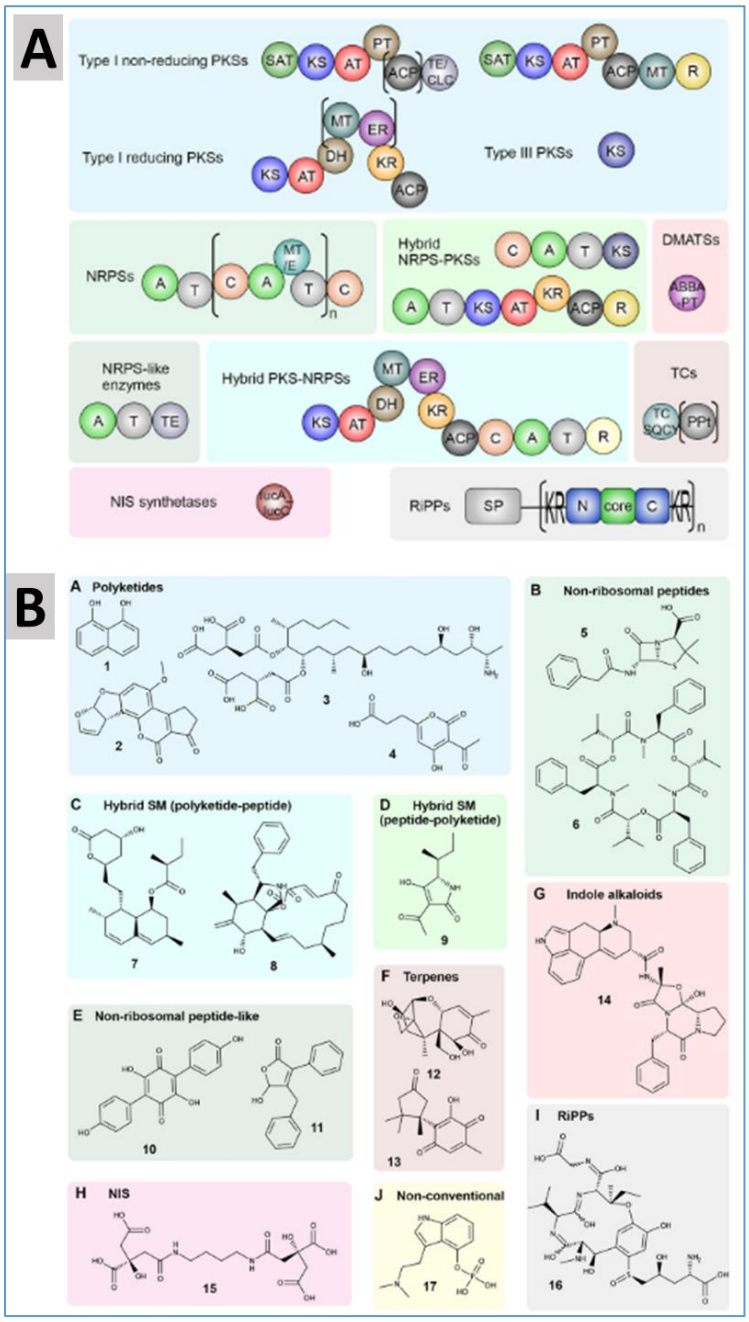


Figure 2. Different core enzymes produce a large variety of secondary metabolites (SMs). A) Different core enzymes are involved in the production of the different classes of SMs. The module organization is shown with the arrangement of conserved domains. SAT: Starter Acyl-carrier protein Trans acylase; KS: Keto-Synthase: AT: AcylTransferase; PT: Product Template; ACP: Acyl-Carrier Protein; TE: ThioEsterase; CLC: Claisen Cyclase; MT: Methyl Transferase; R: Reductase; DH: Dehydratase; ER; Enoyl Reductase; KR: Keto Reductase; A: Adenylation; T: Thiolation; C: Condensation; E: Epimerization; TC: Terpene Cyclase; SQCY: Squalene Cyclase; PPt: PolyPrenyl transferase; ABBA-PT: Aromatic Prenyl Transferase; SP: Signal Peptide. B) Diversity of known fungal SMs. Figure adapted from [19].

The most abundant class of fungal secondary metabolites: polyketides

Polyketides are considered as one of the most important and abundant classes of SMs and their core enzymes consist of several conserved domains [23]. PKSs in general are classified in three major classes: the type I, II, and III PKSs, of which only type I, and type III are commonly found in fungi. Type I PKSs are large enzymes with multiple functional domains that iteratively elongate the backbone from mostly acetyl- and malonyl-CoA. Based on their reductive behavior, these PKSs are further classified in non-reducing PKSs (nrPKS), partially reducing (prPKS), and highly reducing PKS (hrPKS). hrPKSs and prPKSs contain one or more reducing domains that can have a keto reducing, dehydrating, or enoyl-reducing activity [23], [26]. While type I PKSs are well studied, multidomain mega-enzymes with a large variety of produced SMs, much less is known about fungal type III PKSs, that only consist of a single keto-synthase domain and are less abundant in fungi. Type III PKSs catalyze the iterative condensation of a starter fatty acyl-CoA and several extender units. The wide range of starter units they accept combined with the different enzymatic reactions lead to a high variety of products synthesized by type III PKSs [27], [28].

2.2. Prediction, identification, selection, and expression of biosynthetic gene clusters

Since the rise of genomics, it has become possible to study BGCs *in silico*. Although many different SMs have already been discovered, genomics have revealed that there are many more BGCs in the fungal genomes than the number of compounds that are discovered (Fig. 3) [27]. Genomics allows for identification of the BGC and prediction of the class of the SM.

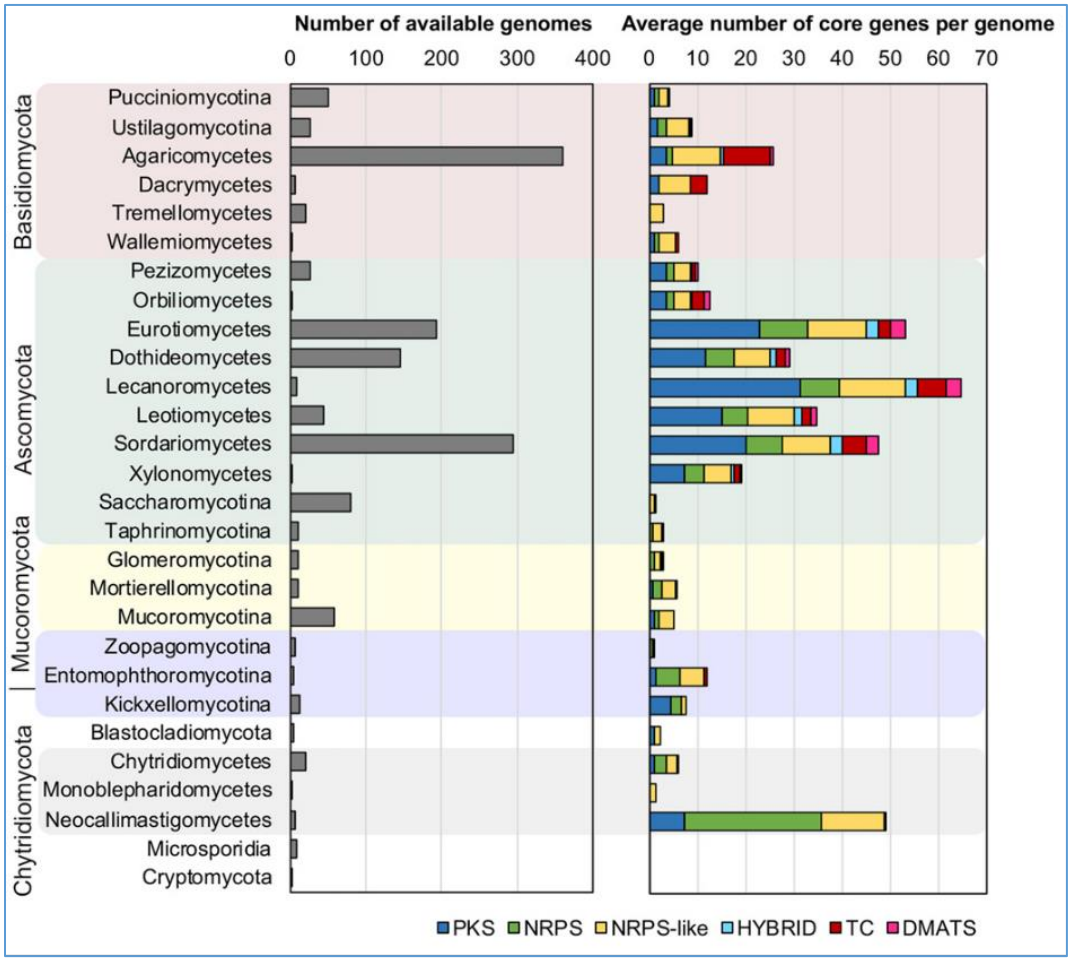


Figure 3. Number of available fungal genomes and the abundance of biosynthetic gene clusters in fungal genomes per fungal class. Figure retrieved from [19].

Several tools have been developed that enable the identification and annotation of BGCs [29]–[32]. As the predicted amount of fungal BGCs is large, it is necessary to prioritize BGCs for characterization. To prevent rediscovery of secondary metabolites and to select for BGCs that likely produce novel compounds or variations to known compounds, several strategies can be used for BGC prioritization. Such strategies include phylogenetic analyses, comparative genomics, and BGC selection based on genes coding for enzymes with specific enzymatic functions in the BGC [12], [19], [31], [33]–[36].

After selection of the BGCs, functional characterization of the synthesized compounds of the BGC is needed. The methods required for this depend on many different aspects [31], [37]. If the fungus is cultivatable and the BGC is not expressed during standard conditions, the BGC is referred to as a silent BGC. These silent BGCs also explain the big difference in encoded pathways and known compounds, as the BGCs must be stimulated to release the SM. Therefore, silent BGCs are also more likely to produce uncharacterized molecules. Modifications of culture conditions, known as the OSMAC (One Strain, Many Compounds) method, or co-cultivation with other organisms can be applied to stimulate expression of the silent BGC. If this does not activate the pathway, genetic modifications can be applied. As BGCs can be regulated by BGC-specific transcription factors, introducing an extra copy of such a transcription factor under control of an

inducible or constitutive promoter can activate the BGC [19], [31], [38]. However, if the BGC is not present in a tractable fungus or the previously mentioned techniques do not yield any SMs, heterologous expression strategies can be applied [31]. During heterologous expression, one or more genes or even a whole BGC of interest from a donor species are introduced into the heterologous host. Several bacterial, yeast, and filamentous fungus species are used as host for heterologous expression. The filamentous fungus *Aspergillus oryzae* NSAR1 strain is, for example, widely used as host organism. *A. oryzae* has been used for over 2,000 years in the Japanese fermentation industry to produce foods, such as sake and soy sauce. This fermentation process is based on the abundant production of taka-amylase (a-1,4-glucan-4-glucanhydrolase). The *amy*B gene located in the *amy*B expression cassette is responsible for this production and is induced when malto-oligosaccharides are available, and repressed when glycerol and glucose are present [39]. This feature is used in the heterologous expression platform using *A .oryzae* M 2-3 used for tenellin pathway investigation [39]. For heterologous expression, the *A. oryzae* strain NSAR1 is often used because of several reasons. For one, heterologous expression of a BGC originating from a donor filamentous fungus may be more successful in a filamentous fungal host than in a non-filamentous fungus. Secondly, the strain produces secondary metabolites to limited extend of its own under normal culture conditions. Furthermore, *A. oryzae* has the ability to secrete large amounts of proteins and the NSAR1 strain enables multiple rounds of transformations, because it is a quadruple auxotroph that is not able to synthesize arginine, methionine, adenine, and not able to express the nitrate reductase gene [31], [40], [41], [39].

For heterologous expression to be successful, the genetic sequence must be correctly cloned or synthesized, and the gene should start and stop at the right positions. Additionally, as correct intron splicing cannot be guaranteed in all hosts, including *A. oryzae* NSAR1, the introns should be predicted accurately and precisely, to be able to only introduce the coding DNA sequence of the genes [39], [41]. Unfortunately, the tools that are currently used to predict genes and gene models are still prone to errors and curation of genes is therefore necessary to correct wrongly predicted gene models [19], [42], [43].

Research objectives

The biosynthetic potential of fungal BGCs has been greatly underappreciated for a long time. However, fungal genomes can be a great source for novel SMs, for example as antimicrobials or pesticides, which are both needed due to the rise of antimicrobial and pesticide resistance, respectively. To exploit the fungal genomes for novel SMs, active development of bioinformatic tools to more accurately predict and correct the gene models in BGCs is needed. Heterologous expression of the predicted BGCs is a promising technique for production of novel compounds. In this project, development of a new bioinformatic tool is launched to enable automated curation of gene models *in silico*, and in parallel, two BGCs are heterologously expressed to gain more knowledge on these unknown SMs.

Since the tools that are currently predicting fungal gene models are still prone to errors, manual curation is needed to correct them. With the increasing number of available fungal genomes and accessible BGCs every day, this is very labor intensive. An automated tool that accurately performs quality checks on the predicted gene models and corrects the errors, would decrease the number of gene models that require manual curation.

Furthermore, previously, two potentially novel biosynthetic pathways were deduced from studies of polyketides in the lichen forming *Ramalina peruviana* (Lecanoromycetes) and the entomopathogenic *Beauveria bassiana* (Sordariomycetes) (FNP lab, unpublished data) (Supplementary materials, Table S1). The class of Lecanoromycetes is the largest class of lichenized fungi [44]. In the light of limited access to strains and given that lichen mycobionts growth is very limited and does not allow genetic manipulations with parental strains, heterologous expression was chosen as a strategy to get access to both pathways' products. An ortholog of the BGC found in *R. peruviana* is present in the genome of *A. melleus* [RETRACTED], designated as cluster [RETRACTED] in accordance with antiSMASH prediction (Fig. 4). The core gene of this BGC encodes for a type I nrPKS, unrelated to another BGC with experimentally verified known SM product [45]. This makes this BGC very interesting as it might produce an unknown SM. Although recently heterologous expression of a BGC from a lichen-forming fungus in a non-lichenizing fungus was successfully performed [46], [47], heterologous expression in non-lichen hosts and functional characterization of BGCs originating from Lecanoromycetes has proven difficult in the past for unknown reasons [48]. Orthologs of Lecanoromycete BGCs can produce the same compounds as the original Lecanoromycete BGCs [34]. Therefore, an ortholog of this promising BGC was chosen for elucidation. *A. melleus* was chosen for this purpose, because it is easily cultivatable, is present in the Westerdijk Institute collection, and there is an assembled genome of good quality available, which allows for amplification or synthesis of the genes present in the BGC [45].

The second chosen BGC is present in *B. bassiana* [RETRACTED], under the name cluster [RETRACTED]. *B. bassiana* is an entomopathogenic and endophytic filamentous fungus [49]. The fungus is widely used in Thailand as biocontrol agent in the agriculture and is studied for its secondary metabolite production [50], [51]. In previous research performed by the Fungal Natural Products group at the Westerdijk Institute, in collaboration with a research group of the BIOTEC research Institute (NSTDA, Thailand), cluster [RETRACTED] was selected for functional characterization. This cluster showed expression during very early stages of insect infection and

was not linked to any known or characterized SM. Therefore, this cluster was selected for functional characterization of its SM and its possible role in entomopathogenicity [45].

[FIGURE RETRACTED]

Figure 4. Phylogenetic tree of *Ramalina peruviana* cluster [RETRACTED]. The blue box indicates the biosynthetic pathway of interest, *Ramalina peruviana*, cluster [RETRACTED]. The red box indicates the homologous biosynthetic pathway *Aspergillus melleus* [RETRACTED] cluster [RETRACTED], which is used for functional characterization of the biosynthetic compound.

Materials and methods

Biological strains and cultivation

Bacterial strains and cultivation

Chemically competent bacterial cells were used for bacterial transformations (Table 1). For plasmids that include the *ccdB* gene, coding for the toxic CcdB protein, One Shot™ *ccdB* Survival™ 2 T1 *Escherichia coli* cells (Thermofisher, Carlsbad, CA, USA) were used for transformations. JM109 *E. coli* cells were used for transformation after pGEM®-T Easy reactions and to prepare seamless ligation cloning extract (SLiCE) cell extracts. For other plasmids, DH5*α* cells were used (Thermofisher, Carlsbad, CA, USA). In liquid culture, the cells were grown overnight in lysogeny broth (LB) (Supplementary materials, Table S2) at 37 °C with shaking at 200 rpm, with adequate selection depending on the plasmid backbone (Table 2). Strains were stored in 17% glycerol at -80 °C.

For disk diffusion assays, *E. coli* strain DH5*α* and *Bacillus subtilis* cells were used. Both were grown on lysogeny broth agar (LA) (Supplementary materials, Table S2) plates at 37 °C for one night and stored at 4 °C on plates until further use.

Table 1. Different *Escherichia coli* strains and their corresponding genotype.

Cells	Genotype
DH5*α*	F- Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(rk-, mk+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 *λ*-
2 T1	F-*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*Δ139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G *fhu*A::*IS2*
JM109	*end*A1, *rec*A1, *gyr*A96, *thi*, *hsd*R17 (rk–, mk+), *rel*A1, *sup*E44, Δ(*lacpro*AB), [F′ *tra*D36, *pro*AB, *laq*IqZΔM15]

Table 2. Selection antibiotic for the backbone plasmids *Escherichia coli* was transformed with.

Plasmid backbone	Selection in *E. coli*
pTYGSade, pTYGSarg	50 mg/mL ampicillin
pGEM-T easy	50 mg/mL ampicillin
pEYA2	35 mg/mL chloramphenicol
pDNOR207	25 μg/mL gentamicin
pTWIST	50 μg/mL kanamycin

Fungal strains and cultivation

The diploid and uracil auxotrophic *Saccharomyces cerevisiae* strain Δura3 was used for yeast transformations to enable homologous recombination of genes into entry or destination vectors. Untransformed cells were grown on yeast extract peptone dextrose (YPD) agar (Supplementary materials, Table S2) at 30 °C for 1-2 days and stored at 4 °C. For maintenance of plasmids in the cells after transformation, cells were grown on synthetic dropout media plates SDM JC (Supplementary materials, Table S2) or in liquid culture in SDM JC without select agar at 30 °C, 200 rpm for 12-16 hours.

For transformation of *Aspergillus oryzae* with the expression vectors, *A. oryzae* strain NSAR1 was used (argB−, niaD−, sC− and adeA−). *A. oryzae* was grown on malt extract agar (MEA) plates (Supplementary materials, Table S2) for 4-6 days at 28 °C, until spores were formed. Plates were stored at 15 °C.

For disk diffusion assays, *Candida albicans* CBS562 and *Penicillium rubens* were used. *P. rubens* was cultured on MEA plates at room temperature and stored at room temperature until further treatment. *C. albicans* was kindly provided by Auke de Jong, MSc, Westerdijk Institute.

Nucleic acids methods

Polymerase chain reaction

To obtain genes of the selected BGCs, genes were amplified using polymerase chain reactions (PCR) using Phusion™ High-Fidelity DNA polymerase (Thermofisher Scientific, Wilmington DE), according to the manufacturer's protocol and with annealing of the primers at 60 °C for 1 minute. The primers used were designed by the FNP group at the Westerdijk Institute (Supplementary materials, Table S3). PCR products were loaded onto an 0.8% agarose gel and recovered from gel or from the PCR mix, after digestion of the template plasmid using *DpnI* (Promega, Madison, USA), according to the manufacturer's protocol with 10 units enzyme for 1 hour at 37 °C and inactivation for 20 minutes at 65 °C. After purification, the PCR products were checked by sequencing (Macrogen, Seoul).

Plasmid extraction and DNA purification

To obtain the plasmids for cloning and transformation from *E. coli* strains, Illustra™ plasmidPrep Mini Spin Kit (Cytiva, Marlborough, MA, USA) was used for plasmid extraction according manufacturer's recommendations to perform miniprep. Different from the protocol, plasmids were eluted in 30 µL nuclease-free H_2O and then stored at -20 °C. Midiprep was performed using the Nucleobond® Xtra Midi Kit (Macherey-Nagel) or the PureLink™ HiPure Plasmid Midiprep Kit (Invitrogen) according to the manufacturer's protocol. To extract plasmids from *S. cerevisiae*, Zymoprep™ Yeast Plasmid Miniprep II (ZYMO RESEARCH, Orange, CA, USA) was used according to the manufacturer's protocol. These plasmids were eluted in 10 µL nuclease-free H_2O and stored at -20 °C. Concentrations of the plasmids were quantified with Nanodrop 1000-spectrophotometer (Thermofisher Scientific, Wilmington DE) and correct integration of the gene into the plasmids was verified by sequencing (Macrogen, Seoul).

PCR products of the amplified genes of the selected BGCs were purified using the Geneclean® II Kit (MPbio, Solon, OH, USA) for PCR products that are later used in yeast transformation, or the NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel) for PCR products used for SLICE or BP Gateway®-cloning (Invitrogen) according to the manufacturer's protocol.

Products purified with the NucleoSpin® Gel and PCR Clean-up Kit were quantified with Nanodrop 1000-spectrophotometer (Thermofisher Scientific, Wilmington DE), whereas products purified with the Geneclean® II Kit (MPbio, Solon, OH, USA) could not be quantified this way, due to residual silica particles that influence the quantification using the Nanodrop 1000-spectrophotometer (Thermofisher Scientific, Wilmington DE).

Calf intestinal alkaline phosphatase treatment and ethanol precipitation

Calf intestinal alkaline phosphatase (CIAP) treatment is used in this study to remove the 5′ phosphate groups from linearized plasmids, to prevent self-ligation. CIAP reaction mixtures and CIAP stop buffer were made according to the manufacturer's protocol (CIAP, Promega). The reaction mixture was incubated at 37 °C for 1 hour and hereafter inactivated by addition of 300 µL CIAP stop buffer. 1 Volume of phenol:chloroform:isoamyl alcohol (25:24:1) (Invitrogen) was added and the mixture was vortexed and then centrifuged for 5 minutes, 14,000 rpm. The aqueous phase was transferred to a new tube and 0.5 volumes of 3 M sodium acetate, and 2 volumes of 100% ethanol were added. After 15 minutes incubation on ice of the mixture, the tubes were centrifuged at maximum speed for 30 minutes. The supernatant was discarded, and the pellet was washed with 70% ethanol and centrifuged at maximum speed for 15 minutes, after which the supernatant was discarded. The pellet was dried for 30 minutes to air and then dissolved in 25 µL nuclease-free H_2O. The DNA yield was quantified using the Nanodrop 1000- spectrophotometer (Thermofisher Scientific, Wilmington DE).

Plasmid construction

pGEM®-T Easy vector

The *B. bassiana* tailoring gene *Aminotransferase class I and II* ([RETRACTED], [RETRACTED]) (*AMI*) was ordered synthetically (Twist Bioscience, CA, USA) and cloned into the pGEM®-T Easy vector (Promega, Madison WI) according to the manufacturer's protocol to enable conservation of the gene in *E. coli* at -80 °C**.** After *E. coli* JM109 transformation, white colonies were selected and correct insertion of the gene was checked by digestion of 1 µg plasmid (pGEM-T::BeaBas_*AMI*) with 5 units *NotI* (Promega, Madison, USA) for 1-2 hours at 37 °C. The digestion mix was loaded onto 0.8% agarose gel and would give two bands (2981 bp and 1429 bp) if the gene was correctly inserted.

Seamless ligation cloning extract (SLiCE)

To clone single fragments into a vector, a modification of the SLiCE method was used [52]. The SLiCE method is an efficient, easy, quick, and inexpensive cloning method that ligates a fragment in a vector backbone using homologous recombination that occurs in the cell extracts of *E. coli*.

Seamless ligation cloning extract (SLiCE) preparation

For preparation of the SLiCE *E. coli* cell extracts, strain JM109 was used. JM109 cells were inoculated in 5 mL LB for overnight incubation. 1 mL of this culture was transferred to 50 mL LB (Supplementary materials, Table S2) in a 250 mL flask and incubated at 37 °C, 200 rpm until the OD600 was 2.0-3.0. The culture was centrifuged for 10 minutes at 4 °C, 5,000 rcf. The supernatant was discarded, and the pellet was washed with 50 mL ice cold MilliQ H_2O. The suspension was centrifuged again for 10 minutes, 4 °C, 5,000 rcf, and the supernatant was discarded. The pellet was resuspended in 1.2 mL buffered cell lytic buffer B lysis reagent (10% CellLytic™ B Cell Lysis reagent (Sigma-Aldrich), 50 mM Tris HCl pH 7.5, 200 mM NaCl; 0.22 µm filter sterilized), incubated for 10 minutes at room temperature, and then centrifuged for 2 minutes at 4 °C, 20,000 rcf. The supernatant was transferred to a fresh tube on ice and mixed with 1 volume 87% sterile glycerol. The mixture was distributed in 40 µL aliquots, frozen in liquid nitrogen and stored at -80 °C.

Seamless ligation cloning extract (SLiCE) reaction

SLiCE *E. coli* cell extracts were defrosted on ice and SLiCE 10 x buffer (0.5 ml/ml 1 M Tris-HCl pH 7.5, 20.3 mg/ml $MgCl_2$*$6H_2O$, 5 mg/ml ATP, 1.54 mg/ml dithiothreitol (DTT); 0.22 μm filter sterilized) was defrosted at room temperature. 1 μL cell extract was mixed with 1 μL SLiCE 10 x buffer, 100 ng linearized and CIAP-treated plasmid, and 200 ng DNA fragments. Nuclease-free H_2O was added up to a final volume of 10 μL. This mixture was incubated at 37 °C for 30 minutes, and 10 μL of the mixture was used to transform *E. coli* cells.

Construction entry and destination vectors for LR Gateway® cloning

Aspergillus melleus [RETRACTED] *cluster* [RETRACTED]

For *A. melleus* [RETRACTED] cluster [RETRACTED] (AspMel), the non-reducing *PKS* (*AME1*) ([RETRACTED], [RETRACTED]) was ordered synthetically. A large sequence of the 5' side of *AME1* was in a pTWIST plasmid (Twist Bioscience, CA, USA) ([RETRACTED]) and the other part ([RETRACTED]) was synthesized as gene fragments. SLiCE was used to create the plasmid containing the complete *AME1* (pTWIST::AspMel_*AME1*).

Tailoring genes *FAD binding protein* (*FAD*) ([RETRACTED], [RETRACTED]), *O-methyltransferase* (*MET*) ([RETRACTED], [RETRACTED]), and *questin oxidase-like* (*QUEST*) ([RETRACTED], [RETRACTED]) were ordered synthetically in pTWIST plasmids (Twist Bioscience, CA, USA), which were used as template for PCR. The last tailoring gene for this BGC, *enoyl-(acyl carrier protein) reductase* (*ER*) ([RETRACTED], [RETRACTED]) was amplified using the *A. melleus* gDNA as template [45].

For *FAD*, *MET*, and *QUEST*, yeast homologous recombination (YHR) was used to create the LR destination vector, pTYGSarg::*F*:*M*:*Q*. 1 μg of pTYGSarg was digested using 10 U *SgsI* (*AscI*) (ThermoFischer Scientific) according to the manufacturer's protocol for 1 hour at 37 °C, and inactivated for 20 minutes at 65 °C. The digested plasmid was verified on 0.8% agarose gel and purified from the digestion mixture using Geneclean® II Kit (MPbio, Solon, OH, USA). The digested plasmid and the amplified *FAD*, *MET*, and *QUEST* were used in to transform *S. cerevisiae* Δura3 to enable YHR to create the destination vector pTYGSarg::*F*:*M*:*Q*. Amplification of pTYGSarg::*F*:*M*:*Q* was achieved by transformation of One Shot™ *ccdB* Survival™ 2 T1 *E. coli* cells with this destination vector.

ER was cloned into pTYGSade using SLiCE and LR Gateway® cloning (Invitrogen). First, 1 μg pEYA2 was digested with 5 units *NotI* (Promega, Madison) for 1-2 hours at 37 °C and the reaction was inactivated by incubation for 20 minutes at 65 °C. SLiCE was used to clone *ER* into the digested pEYA2 to obtain pEYA2::*ER*, which was amplified in *E. coli* DH5α cells. After extraction of pEYA2::*ER* from *E. coli*, LR Gateway® cloning (Invitrogen) was done according to the manufacturer's protocol with 70 ng pEYA2::*ER* as LR entry vector and 150 ng pTYGSade as LR destination vector to obtain pTYGSade::*ER*, which was transformed into *E. coli* DH5α cells for plasmid selection and amplification.

Beauveria bassiana [RETRACTED] *cluster* [RETRACTED]

For *Beauveria bassiana* [RETRACTED] cluster [RETRACTED] (BeaBas), the reducing *PKS* (*BBT1*) ([RETRACTED], [RETRACTED]) was ordered in two fragments of the same size in pTWIST plasmids (Twist Bioscience, CA, USA). Those fragments were amplified using PCR and used with

pEYA2 digested with *NotI*, as described before, to transform *S. cerevisiae* Δura3 and enable YHR to obtain pEYA2::BeaBas_*BBT1*. pEYA2::BeaBas_*BBT1* was amplified in *E. coli* DH5α cells and used as entry vector in an LR Gateway® Cloning (Invitrogen) reaction with pTYGSarg as destination vector, according to manufacturer's protocol, to obtain expression vector pTYGSarg::BeaBas_*BBT1*.

The only BeaBas tailoring gene *AMI* was amplified from pGEM-T::BeaBas_*AMI* by PCR using primers designed to be compatible with BP Gateway® Cloning (Invitrogen), according to the manufacturer's protocol. A BP Gateway® Cloning reaction was done using this PCR product and donor vector pDONR207 according to protocol and *E. coli* DH5α cells were transformed with the final product: pDONR207::BeaBas_*AMI*.

Transformations

Transformation of *Aspergillus oryzae* NSAR1 with expression vectors of the biosynthetic gene clusters

Conidia from sporulating *A. oryzae* NSAR1 plates were suspended in 5 mL sterile MilliQ. 50 mL malt extract broth (MB) (Supplementary materials, Table S2) in a 250 mL Erlenmeyer flask was inoculated by adding 1 mL spore suspension, and was incubated overnight at 28 °C with shaking at 200 rpm. The next day, transformation was performed according to [53] protocol. Arginine or adenine selection was used for transformants conferring plasmids with backbone consisting of pTYGSarg or pTYGSade, respectively. After transformation, the plates were incubated at 30 °C for 7 days. Grown transformants were then individually plated onto selection plates and yeast malt agar (YMA) (Supplementary materials, Table S2).

Transformation *Saccharomyces cerevisiae* Δura3 to enable yeast homologous recombination

Diploid *S. cerevisiae* strain with *URA3* auxotrophic marker was used for transformation. The protocol was adapted from [54]. In brief, *S. cerevisiae* was plated from the -80 °C glycerol stock onto YPD plates and incubated at 30 °C for 2 days. Using a sterile loop, 5 mL of 5% YPD was inoculated with yeast cells and incubated for 16 hours at 30 °C, 200 rpm. 2.5 * 10^8 cells were then transferred to 50 mL 5% YPD and incubated for 4 fours at 30 °C, 200 rpm. The cells were harvested by 5 minutes centrifugation at 2,000 rpm, washed once with 10 mL sterile MilliQ, pelleted again, and this pellet was reconstituted in 300 µL sterile MilliQ. 50 µL of this suspension was carefully mixed with 250 µL sterile DTT (100 mM) and incubated at room temperature for 10 minutes, after which it was centrifuged for 15 seconds at 15,000 rcf and the supernatant was discarded. The pellet was reconstituted with 500 µL PLTE solution (for 1 ml: 800 µL 50% PEG 4000 (Sigma-Aldrich), 100 µL 1 M LiAc, 20 µL 50 mM EDTA, 10 µL 1M Tris HCL pH 7.5 , 70 µL H_2O), 4 µL of each DNA fragment, and 50 µL denatured salmon semen (2 mg/mL). This was incubated for 1 hour at 30 °C, inverting the tube every 20 minutes. Samples were incubated for 15 minutes at 45 °C and immediately spun for 15 seconds at 15,000 rcf. The cells were washed once with 200 µL sterile MilliQ, centrifuged for 15 seconds 15,000 rcf and the supernatant was discarded. The pellet was resuspended in 1 mL 5% YPD and incubated at 30 °C for 30 minutes. The cells were harvested by 15 seconds centrifugation at 15,000 rcf, reconstituted in 100 µL sterile MilliQ, and directly plated

onto synthetic dropout media plates, SDM JC. The plates were then incubated at 30 °C for 4 days. Grown transformants were plated onto fresh SDM JC plates and incubated for 3 days at 30 °C.

Transformation *Escherichia coli*

Heat shock competent *E. coli* cells were used to perform transformation of *E. coli* strains. 10 µL of LR, BP, or SLICE reaction, or 10 ng DNA was added to 50 µL cells and incubated for 30 minutes on ice. The mixture was then incubated for 45 seconds at 42 °C and put directly back on ice for 2 minutes. 950 µL LB was added, and the cells were incubated at 37 °C, 200 rpm for 1 hour. The cells were pelleted by spinning down for 15 seconds, 15,000 rcf. The supernatant was discarded, cells were resuspended in 100 µL LB, immediately plated onto plates with selection medium and incubated at 37 °C overnight. Grown colonies were plated onto fresh selection medium and incubated overnight at 37 °C. Grown colonies were then verified by colony PCR or digestion analysis. 5 or 50 mL LB with antibiotic selection was inoculated with the colony and incubated overnight at 37 °C, 200 rpm, for a miniprep or midiprep, respectively.

Metabolite production and characterization

Secondary metabolite production by transformed *Aspergillus oryzae* NSAR1

To allow *A. oryzae* NSAR1 transformed with genes from the BGC to produce metabolites, 50 mL YMA + starch (Supplementary materials, Table S2) was inoculated by addition of a piece of agar of approximately 1 cm^2 with mycelium of the transformant. This was incubated for 4-5 days at 28 °C shaking at 200 rpm.

Metabolite extraction from media

To extract the secreted fungal metabolites from the medium, the fungal biomass was filtered out using miracloth (Merck). Secondary metabolites were extracted by addition of 1 volume ethyl acetate to the medium and placed on an orbital shaker at room temperature for two hours. Hereafter, the samples were spun down for 5 minutes at maximum speed and the organic phase was transferred to a new vial. Samples originating from these organic phases are hereafter called "first extraction". The remaining water phase was acidified to pH ~ 3.0 and subjected to a second round of extraction with 1 volume ethyl acetate, one night orbital shaking, and the organic phase was transferred to a new tube again, hereafter called "second extraction". The first and second extraction samples were evaporated under the nitrogen flow, and the resulting solids were reconstituted in 400 µL acetonitrile, resulting in two separate extracts per sample: the first and second extraction.

Characterization of the biosynthetic gene cluster products

For characterization of biological activity of the produced compounds, 75 µL of the crude extracts was placed onto sterile 9 mm antibiotic assay disks (Machery-Nagel), which were dried to air under sterile conditions. For positive control, 75 µL LB with ampicillin was used for bacteria, or 75 µL amphotericin B for fungal species. As negative control, 75 µL acetonitrile was used. Additionally, 200 µL and 100 µL of two different concentrations newly prepared amphotericin B (2.5 µg/mL and 0.25 µg/mL) were included to test the toxicity of amphotericin B against *C. albicans*.

To study antibiotic activity, 500 µL of *E. coli* and *B. subtilis* resuspended in sterile H_2O was plated onto separate LA plates, dried to the air, and the dried disks were placed upside down

onto the bacteria on agar. To study antifungal activity, 500 µL in sterile H_2O resuspended spores from *P. rubens* was plated onto MEA plates and 550 µL of resuspended *C. albicans* cells were plated onto Sabouraud glucose agar (Supplementary materials, Table S2). Plates were also dried to the air and the disks with crude extracts were placed upside down on the plates with fungi. The plates with *P. rubens* were incubated at room temperature for 3 days, and the *C. albicans* plates were incubated at 35 °C for 1 day until evaluating the results.

For chemical analysis of the produced compounds, high-performance liquid chromatography (HPLC) and liquid chromatography-mass spectrometry (LC-MS) analyses of crude extracts were performed on Shimadzu LC-2030 system with Photodiode Array (PDA) detection (190-800 nm) equipped with Shimadzu Shim-pack GIST C18-HP reversed phase column (3 µm, 4.6 × 100 mm). Following protocol was used with 1.0 ml/min flow rate when only HPLC analysis was performed or 0.5 ml/min flow rate when both HPLC and LC-MS analyses were performed: linear gradient of buffer B (0-95%) for 20 min, then 5 min 95% B, then 100% buffer A for 5 min. Water supplied with 0.1% TFA for HPLC or 0.05% formic acid for LC-MS was used as buffer A, and acetonitrile (LC-MS grade) with 0.1% TFA for HPLC or 0.05% formic acid as buffer B.

Development of automated gene curation tool

The automated gene curation tool is written in Python (v. 3.8.5). The tool accepts .gbk files and aligns inserted protein sequences using MAFFT (v. 7.475) [55]. Then, trimAl (v. 1.2rev59) [56] is used to eliminate sequences with less than 40% similarity with the other called amino acids. The code will become available on GitHub in the near future.

Results

To reduce the number of genes that need manual curation in the identified biosynthetic regions, a bioinformatic tool was developed to enable automated gene curation. Additionally, two fungal BGCs were selected to perform functional characterization of the produced compound: *A. melleus* [RETRACTED] cluster [RETRACTED] (AspMel) and *B. bassiana* [RETRACTED] cluster [RETRACTED] (BeaBas).

Automated gene curation tool

There are many tools that can predict genes with high proficiency relying on genetic signatures. However, mistakes are common in these gene prediction tools and to detect these problems, some tools have been developed to automatically improve the annotations of predicted gene models [57]. One such a tool is called OMGene. OMGene tries to improve the annotated gene models by making putative gene models from the input data that holds information on the genome and gene coordinates. These putative gene models are then used to evaluate the exons in the genes and to improve the gene models [57]. Another tool that was developed is based on gene-by-gene analysis to improve and re-annotate gene models [42]. This tool uses the genomic loci encoding for the proteins to align the target and informant genes and uses graph-theory to determine the start, stop, and intron-exon boundaries of the genes [42]. The tool developed in this study is based on gene-by-gene analysis and uses informants, like the ABFGP method [42]. However, it uses a different strategy than the graph-theory used in the ABFGP method to evaluate the exons. Like the OMGene method, a scoring method for the alignment will also be incorporated later in the tool to evaluate the alignment of the adjusted gene models.

The tool to perform automated gene curation was written in Python v3.8.5. The goal of the tool is to evaluate predicted gene models by using homologous gene models as reference, to notice any inconsistencies between the homologues, correct them where possible, and to give a report on the changes the tool made, and which parts might still need additional manual curation. The tool currently requires GenBank files as input from the user and creates so-called exon blocks (EBs) that will be used to evaluate and correct the models in follow-up steps/future development (Fig. 5).

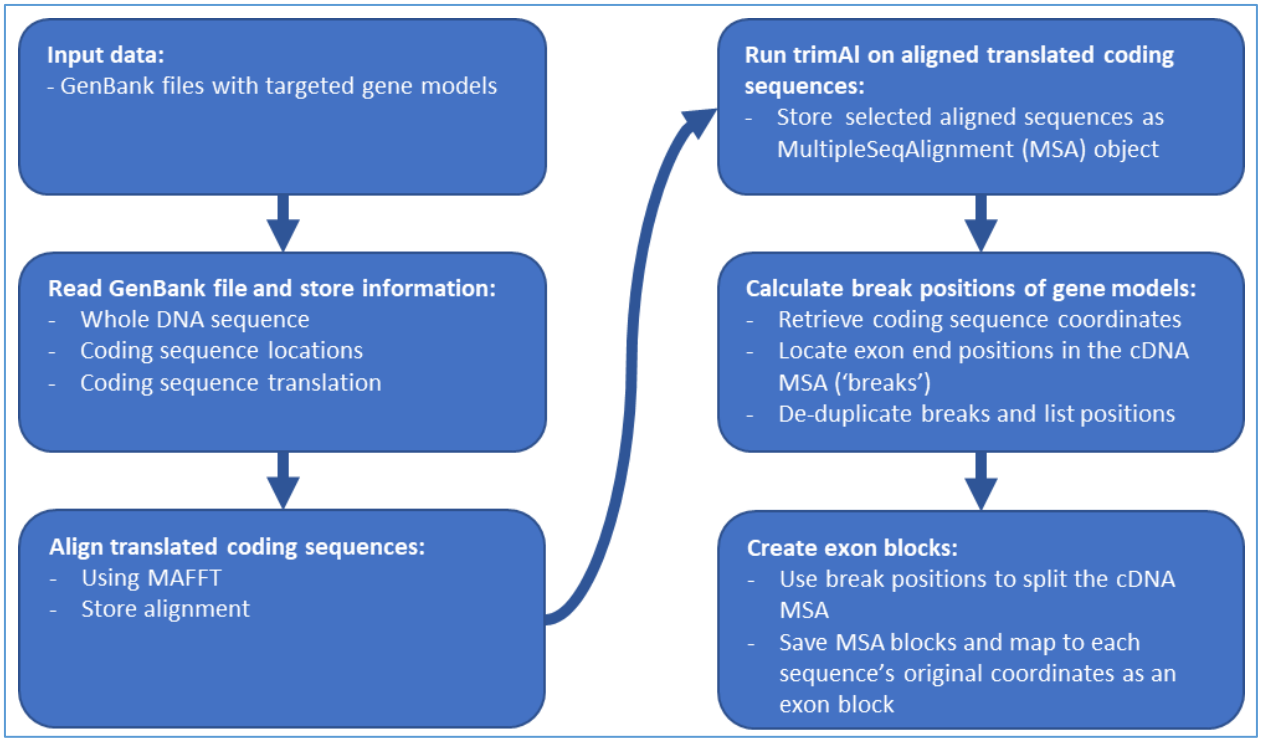


Figure 5. Workflow of the automated gene curation tool.

The tool reads the GenBank files and stores the sequence and information of genes saved in the GenBank file, such as location of exons and the genes' amino acid sequence (AAS). Then, it aligns the AASs of the given genes, using MAFFT v7.475. Hereafter, trimAl 1.2rev59 is used to perform a quality check and to filter out genes with poor alignment quality. For this check, trimAl takes amino acid locations into consideration that have at least a 40% overlap between all the aligned AASs. Proteins that have an AAS that corresponds for less than 40% with the rest of the AASs are filtered out, while the ones with a good alignment are kept in order to use their corresponding gene models as reference. After this filtering step, the AASs are translated back to nucleotide sequences using the original DNA sequences, taking into account the insertions or deletions of amino acids in their multiple sequence alignment (MSA) (Fig. 6).

[FIGURE RETRACTED]

Figure 6. Example of amino acid deletion/insertion and of exon loci. A) Loci of gene models 1-3 in the GenBank sequence. For example, the first exon of gene model 1 starts at the 700th nucleotide and stops at position 742. After the intron (33 nucleotides), the second exon starts at position 775 and the gene model stops at position 805. B) Example of alignment with a triplet nucleotides missing in gene model 2. C) Correspondingly, gene model 2 lacks an amino acid on position 2 compared to gene model 1 and 3, due to a deletion of this amino acid in this gene model.

Hereafter, EBs are made of the selected, aligned gene models (Fig. 6B). The loci that indicate where the exons are located in the genome are modified to fit the back translations of the trimAl selected gene models. This modification results in MSA of cDNA that is linked to information on the location of the start of each of the gene models, the boundaries between introns and exons, and where the gene models end. Based on these locations, the aligned nucleotide sequences are divided into EBs. Each EB holds one aligned exon, or a piece on an exon, of all the

homologous gene models. The first exon block starts at the first nucleotide of the aligned homologues and ends at the first exon-intron border in the list and contains all aligned nucleotides of the homologues until this position (Fig. 7A). The second EB starts after this exon-intron border and stops again at the next exon-intron border and contains all aligned nucleotides of the homologues between these nucleotide positions (Fig. 7B), and so on until the last exon stops (Fig. 7C). This results in one EB if there is no intron in any of the predicted gene model homologues, or several EBs, if introns are present (Fig. 8).

[FIGURE RETRACTED]

Figure 7. Example of exon blocks (EBs). Three gene models divided into the first EB (A), second EB (B), and the last EB (C).

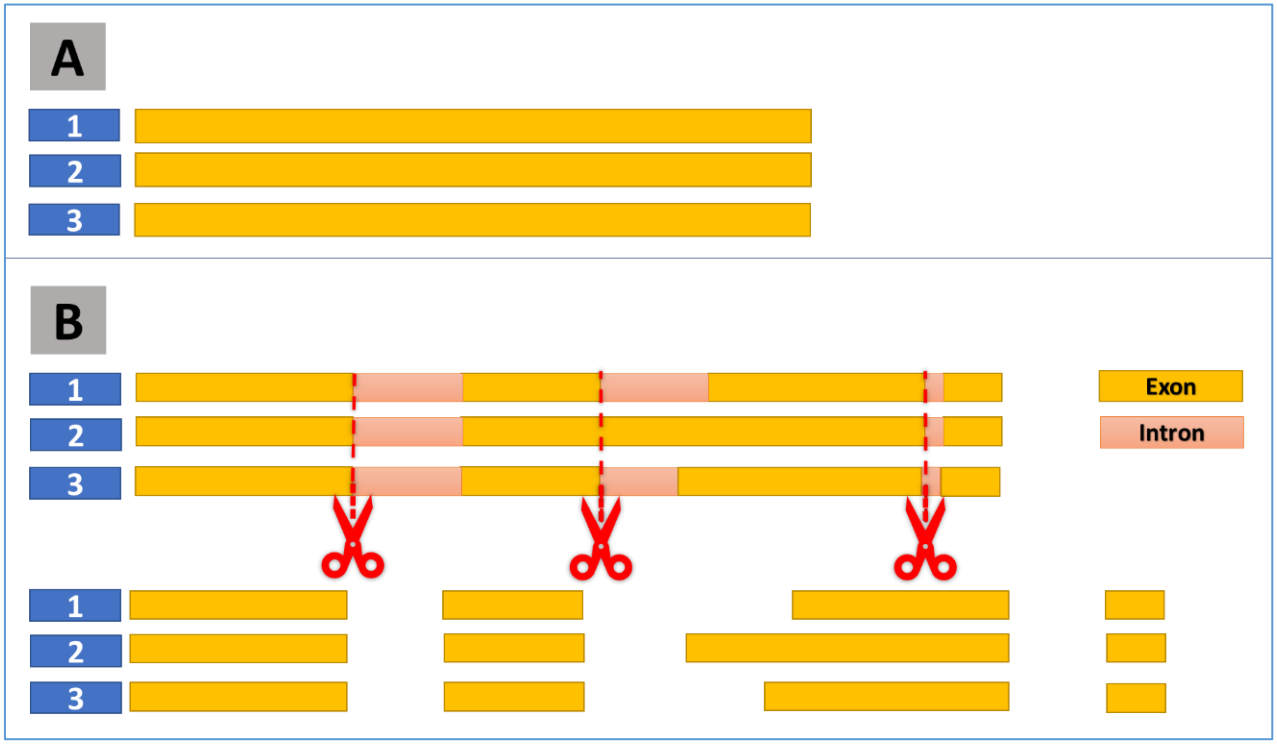


Figure 8. Visualization of exon blocks of three gene models without (A) or with (B) introns.

The EBs are then stored as a MultipleSeqAlignment object from the Biopython library, which is the current state of the script. Further development to evaluate the EBs and provide automated curation is needed, which is discussed in more details in the discussion section.

Heterologous expression of biosynthetic gene clusters

The selected BGCs AspMel and BeaBas are expected to produce novel aromatic and linear polyketides, respectively. The clusters were heterologously expressed in *A. oryzae* NSAR1 and the produced compounds were evaluated on biological activity using disk diffusion assays, and on chemical properties using HPLC and LC-MS.

Construction entry and destination vectors for LR Gateway® cloning

In this research, the tailoring genes (Supplementary materials, Table S1) from AspMel BGC were either amplified from the *A. melleus* genome (*ER*; has no introns) or amplified using PCR from commercially synthesized plasmids (pTWIST) containing the gene without introns (*FAD*, *MET*, *QUEST*) (Fig. 9). *ER* was amplified from the *A. melleus* gDNA and was then cloned into the LR Gateway® entry vector pEYA2 using SLiCE. *FAD*, *MET*, and *QUEST* were cloned into LR Gateway® destination vector pTYGSarg using yeast homologous recombination (YHR). The *PKS* from AspMel BGC (*AME1*) was commercially synthesized as two fragments: 5′ part already cloned into LR Gateway® entry vector pTWIST, and the 3′ part consisting of short gene fragments which were cloned into the pTWIST LR Gateway® entry vector using SLiCE. These cloning events resulted in three plasmids ready for LR Gateway® cloning: pEYA2::AspMel_*ER* and pTWIST::AspMel_*AME1* as entry vectors, and pTYGSarg::*F*:*M*:*Q* as destination vector. LR Gateway® cloning was then performed with these vectors and with empty pTYGSarg and pTYGSade, resulting in expression vectors pTYGSarg::AspMel_*AME1*, pTYGSade::AspMel_*ER*, and pTYGSarg::AspMel_*AME1*::*F*:*M*:*Q* (Fig. 9).

For BeaBas BGC, the *PKS* gene (*BBT1*) was obtained as two equally long fragments cloned into pTWIST plasmids. These fragments (5′ and 3′, respectively) were amplified from pTWIST plasmids using PCR and cloned into LR Gateway® entry vector pEYA2, using YHR. The BeaBas tailoring gene (*AMI*), with a conserved aminotransferase I & II domain, was synthesized and cloned into pGEM®-T easy, from which it was later amplified with PCR using primers with attP flanks. A BP Gateway® cloning reaction was used to clone the gene into LR Gateway® entry vector pDONR207. These cloning events resulted in LR Gateway® entry vectors pEYA2::BeaBas_*BBT1* and pDONR207::BeaBas_*AMI*, which were used in LR Gateway® reactions with pTYGSarg and pTYGSade, respectively, to successfully obtain expression vectors pTYGSarg::BeaBas_*BBT1* and pTYGSade::BeaBas_*AMI* (Fig. 9).

[FIGURE RETRACTED]

Figure 9. Schematic overview of cloning steps to obtain the expression vectors of biosynthetic gene clusters (BGC) *Aspergillus melleus* [RETRACTED] cluster [RETRACTED] (AspMel) and *Beauveria bassiana* [RETRACTED] cluster [RETRACTED] (BeaBas). BGC: biosynthetic gene cluster; nrPKS: non-reducing polyketide synthase; rPKS: reducing polyketide synthase; *ER*: *enoyl reductase*; *F*: *FAD binding domain containing protein*; *M*: *O-methyltransferase*; *Q*: *questin oxidase-like*; PCR: polymerase chain reaction, SLiCE: seamless ligation cloning extract; YHR: yeast homologous recombination.

Transformation of *Aspergillus oryzae* NSAR1 with expression vectors of the biosynthetic gene clusters

To express the BGCs, *A. oryzae* NSAR1 was transformed with the obtained expression vectors and co-transformed with combinations of expression vectors originating from one BGC. The efficiency of the transformation with certain vectors differed (Table 3). Of the co-transformations, only the

co-transformation of pTYGSarg::AspMel_*AME1* with pTYGSade::AspMel_*ER* gained two colonies, while none of the other co-transformations yielded any transformants. Therefore, no transformations of the combinations pTYGSarg::*AME1*::*F*:*M*:*Q* with pTYGSade::*ER*, pTYGSarg::*F*:*M*:*Q* with pTYGSade::*ER*, or pTYGSarg::BeaBas_*BBT1* + pTYGSade::BeaBas_*AMI* are available for analysis in this study. The two clones were still able to grow on selection medium after being replated onto fresh YMA, indicating that they are true transformants with integration of genes in the genome on a position that is compatible with mitotic division (Table 3).

Transformation of *A. oryzae* NSAR1 with a single plasmid was successful with all vectors and resulted in 6 to 11 colonies per transformation plate. However, after a second selection, many of these transformants were not able to grow on selection medium and were therefore excluded from further analysis. An exception on this were the transformants with pTYGSade::BeaBas_*AMI* and those with pTYGSade::AspMel_*ER*. Almost all of these transformants survived the second selection and were likely true transformants (Table 3).

Table 3. Transformation efficiency of the different vectors in *Aspergillus oryzae* NSAR1, and selection of transformants for characterization of produced compounds. The number of colonies after transformation on the selection plate (Number of colonies) and survival of the picked transformants on fresh selection medium (max. 10 picked) after replating on YMA (2nd selection). The notation of picked colonies for liquid culture to obtain compounds for bioactivity and chemical assays are also indicated (Liquid culture) (for results on these assays, see next sections). **Colony 5.2 did not grow during 2nd selection but was inoculated into liquid culture from the first selection medium plate.*

Transformation	Grown colonies	2nd selection	Liquid culture
pTYGSarg::BeaBas_*BBT1*	10	3	1.1 & 1.4
pTYGSarg::BeaBas_*BBT1* + pTYGSade::BeaBas_*AMI*	0	0	
pTYGSade::BeaBas_*AMI*	6	6	3.2 & 3.4 & 3.5
pTYGSarg::AspMel_*AME1* + pTYGSade::AspMel_*ER*	2	2	4.1 & 4.2
pTYGSarg::AspMel_*AME1*::*F*:*M*:*Q*	9	1	5.1 & 5.2*
pTYGSarg::AspMel_*AME1*	9	1	6.1
pTYGSarg::ApsMel_*AME1*::*F*:*M*:*Q* + pTYGSade::AspMel_*ER*	0	0	
pTYGSade::AspMel_*ER*	11	9	8.3 & 8.8 & 8.10
pTYGSade::AspMel_*ER* + pTYGSarg::*F*:*M*:*Q*	0	0	
pTYGSarg::*F*:*M*:*Q*	10	1	10.1

Characterization of the biosynthetic gene cluster products

Transformants were grown on liquid YMA medium containing malt and starch (YMA + starch) to induce PamyB controlled *PKS* expression and obtain secondary metabolites (Table 3). Cultivation of the different transformants showed variations in gained colors of the medium and mycelium after four days growth (Fig. 10). Mycelium of *A. oryzae* NSAR1 transformants with only *FAD*, *MET*, and *QUEST* genes (Fig. 10A) had a slightly darker medium compared to control YMA

+ starch and transformants without any coloring, such as transformants with only *ER* (Fig. 10E), and the mycelia were colored light pink. Transformants with AspMel *AME1* and *FAD*, *MET*, and *QUEST* showed the largest change in color, with dark purple mycelium and purple-red medium (Fig. 10B). *A. oryzae* NSAR1 co-transformed with AspMel *AME1* and *ER* also showed a large change in color, with purple-red medium (Fig. 10C). The mycelium in this samples was also colored purple, but less dark compared to transformants that confer *AME1* with tailoring genes *FAD*, *MET*, and *QUEST*. Mycelium of transformants that confer AspMel *AME1* alone were also colored purple (Fig. 10D), and the media was colored darker than the uncolored transformants (Fig. 10E). The *A. oryzae* transformants with BeaBas *BBT1* or *AMI* did not show any difference in color (data not shown).

[FIGURE RETRACTED]

Figure 10. Media and mycelium of different *Aspergillus oryzae* NSAR1 transformants after 4 days growth in liquid YM at 28 °C, 200 rpm. A) *A. oryzae* transformed with pTYGSarg::*F*:*M*:*Q*. B) *A. oryzae* transformed with pTYGSarg::AspMel_*AME1*::*F*:*M*:*Q*. C) *A. oryzae* transformed with pTYGSarg::ApsMel_*AME1* and pTYGSade::AspMel_*ER*. D) *A. oryzae* transformed with pTYGSarg::AspMel_*AME1*. E) *A. oryzae* transformed with pTYGSade::AspMel_*ER* (F). B) *A. oryzae* NSAR1 transformed with pTYGSarg::*F*:*M*:*Q*.

After extraction of the metabolites from the media, crude extracts were tested for antibiotic activity against *E. coli* DH5α and *B. subtilis*, and on antimycotic activity on *P. rubens,* and *C. albicans* CBS562. Additionally, crude extracts were analyzed using HPLC and LC-MS.

The antibiotic assay revealed that the secondary metabolite produced by *A. oryzae* NSAR1 transformed with AspMel *AME1* and tailoring genes *FAD*, *MET*, and *QUEST* (vector pTYGSarg::AspMel_*AME1*::*F*:*M*:*Q*) inhibited growth of *E. coli* and *B. subtilis* for both the first (Fig. 11) and second extraction of the metabolites in sample 5.1 (data not shown). The other extractions did not show any antibiotic activity.

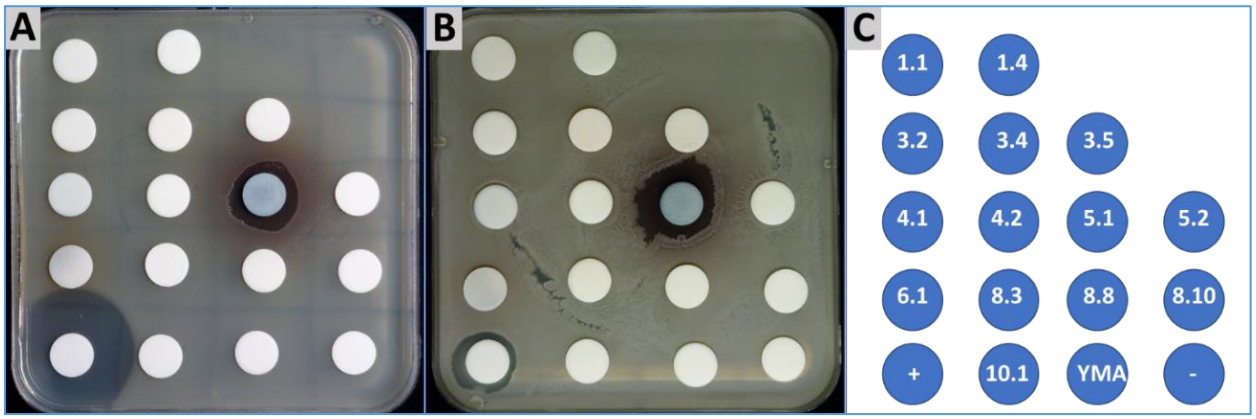


Figure 11. Biological assay of extracted compounds against *Bacillus subtilis* and *Escherichia coli* after 1 day at 37 °C. Disk diffusion assays against A) *B. subtilis*, and B) *E. coli*. C) Overview of extracted compounds on disk, indicating *A. oryzae* transformed with: 1.1 & 1.4: pTYGSarg::BeaBas_*BBT1*; 3.2 & 3.4 & 3.5: pTYGSade::BeaBas_*AMI*; 4.1 & 4.2: pTYGSarg::AspMel_*AME1* and pTYGSade::AspMel_*ER*; 5.1 & 5.2: pTYGSarg::AspMel_*AME1*::*F*:*M*:*Q*; 6.1: pTYGSarg::AspMel_*AME1*, 8.3 & 8.8 & 8.10: pTYGSade::AspMel_*ER*; 10.1: pTYGSarg::*F*:*M*:*Q*; +: positive control (50 mg/mL ampicillin in LB); YMA: extraction from only YMA medium; -: negative control (acetonitrile).

Antimycotic assays were performed against *C. albicans* and *P. rubens.* The disk diffusion assay against *C. albicans* showed that toxicity against *C. albicans* between the first and second

extraction differed, and the positive control was not efficient on the plate with compounds from the second extraction. To test toxicity of different concentrations amphotericin B, an additional disk diffusion assay with different concentrations amphotericin B was included against *C. albicans*, but none of the concentrations showed any growth inhibition (Fig. 12), which should be kept in mind when looking at these results. For the extracted compounds produced by *A. oryzae* NSAR1 transformed with the vectors of the BeaBas BGC, antimycotic activity against *C. albicans* was observed for one sample transformed with pTYGSarg::BeaBas_*BBT1* and all samples transformed with pTYGSade::BeaBas_*AMI*, after the first metabolite extraction (Fig. 13A). *A. oryzae* NSAR1 transformed with pTYGSarg::AspMel_*AME1* and pTYGSade::AspMel_*ER* also showed growth inhibition against *C. albicans*, whereas samples transformed with pTYGSarg::AspMel_*AME1*::*F*:*M*:*Q*, or solely *AME1* or with the genes tailoring *FAD*, *MET*, and *QUEST* (pTYGSarg::AspMel_*AME1* and pTYGSarg::*F*:*M*:*Q*, respectively) had no effect. Of the three extractions from transformants with *ER* (pTYGSade::AspMel_*ER*), two samples inhibited growth of *C. albicans* (Fig. 13A). No compounds from the second extraction inhibited *C. albicans* growth (Fig. 13B). The disk diffusion assay against *P. rubens* showed no toxic effect at all, for none of the extracted metabolites or the positive control (Fig. 14).

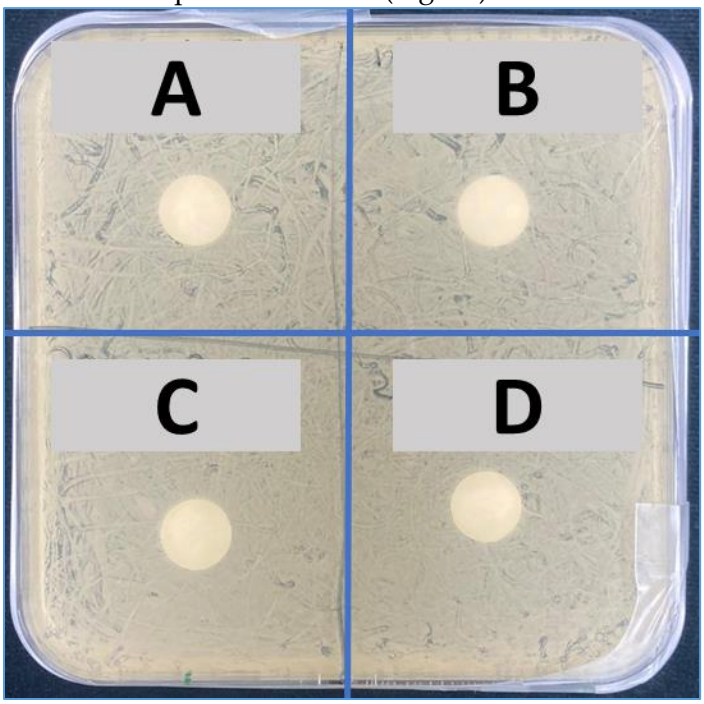


Figure 12. Biological assay of different volumes of different concentrations amphotericin B against *Candida albicans* after 1 day at 35 °C. A) 100 µL, 2.5 µg/mL. B) 200 µL, 2.5 µg/mL. C) 100 µL, 0.25 µg/mL. D) 200 µL 0.25 µg/mL

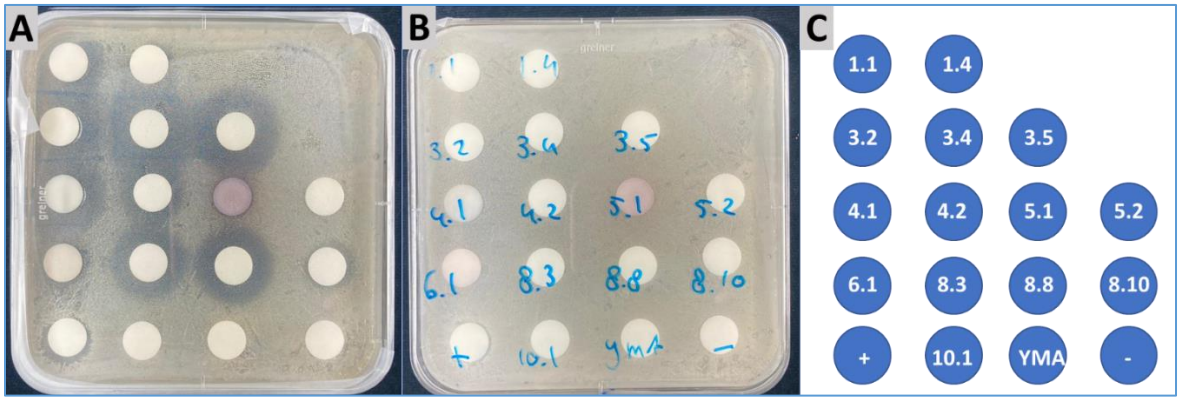


Figure 13. Biological assay of extracted compounds against *Candida albicans* after 1 day at 35 °C. Disk diffusion assays against *C. albicans* using extracts on the disks of A) the first metabolite extraction, and B) the second metabolite extraction. C) Overview of extracted compounds on disk, indicating *A. oryzae* transformed with: 1.1 & 1.4: pTYGSarg::BeaBas_*BBT1*; 3.2 & 3.4 & 3.5: pTYGSade::BeaBas_*AMI*; 4.1 & 4.2: pTYGSarg::AspMel_*AME1* and pTYGSade::AspMel_*ER*; 5.1 & 5.2: pTYGSarg::AspMel_*AME1*::*F*:*M*:*Q*; 6.1: pTYGSarg::AspMel_*AME1*, 8.3 & 8.8 & 8.10: pTYGSade::AspMel_*ER*; 10.1: pTYGSarg::*F*:*M*:*Q*; +: positive control (amphotericin B); YMA: extraction from only YMA medium; -: negative control (acetonitrile).

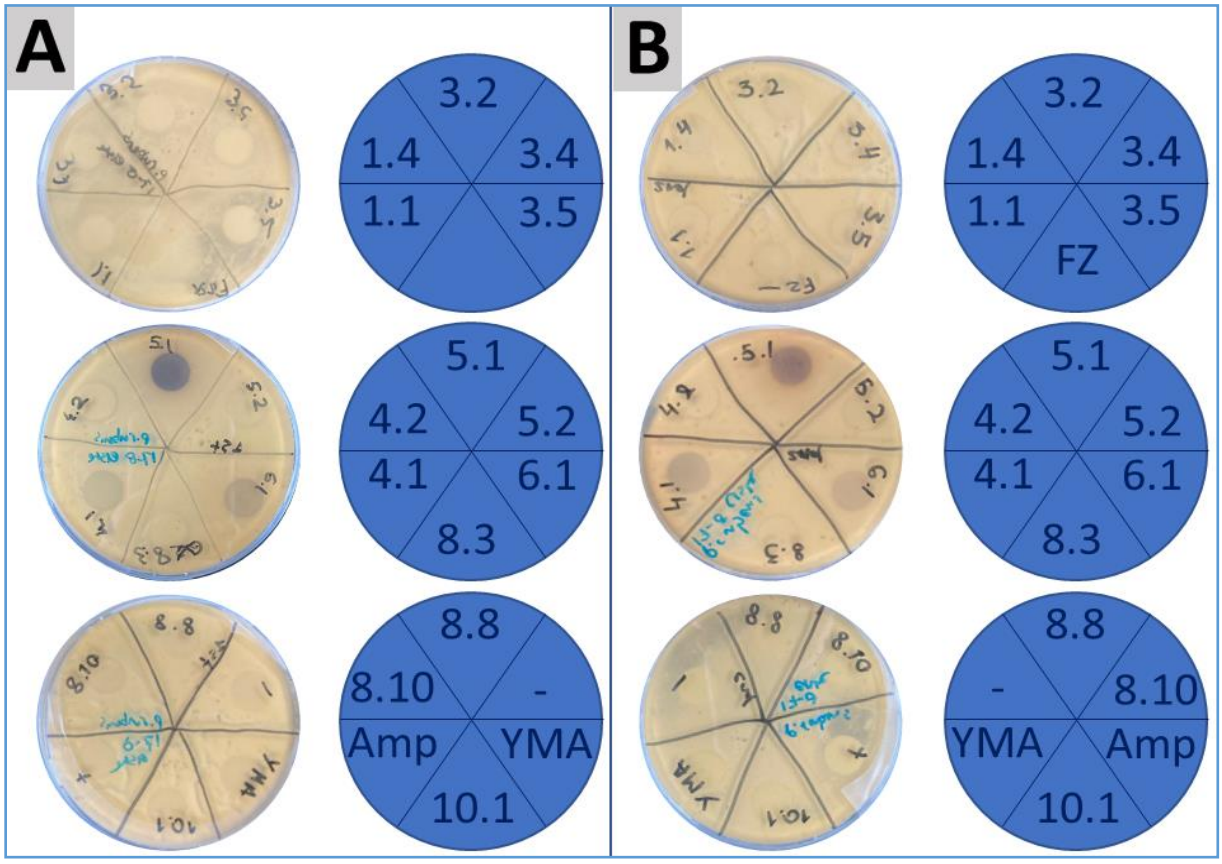


Figure 14. Biological assay of extracted compounds against *Penicillium rubens* after 2 days at room temperature. Disk diffusion assays against *P. rubens* using extracts on the disks of A) the first metabolite extraction, and B) the second metabolite extraction. In blue, an overview of the extracted compounds on disk, indicating *A. oryzae* transformed with: 1.1 & 1.4: pTYGSarg::BeaBas_*BBT1*; 3.2 & 3.4 & 3.5: pTYGSade::BeaBas_*AMI*; 4.1 & 4.2: pTYGSarg::AspMel_*AME1* and pTYGSade::AspMel_*ER*; 5.1 & 5.2: pTYGSarg::AspMel_*AME1*::*F*:*M*:*Q*; 6.1: pTYGSarg::AspMel_*AME1*, 8.3 & 8.8 & 8.10: pTYGSade::AspMel_*ER*; 10.1: pTYGSarg::*F*:*M*:*Q*; Amp: 50 mg/mL ampicillin in LB; FZ: positive control (amphotericin B); YMA: extraction from only YMA medium; -: negative control (acetonitrile).

The HPLC analysis of compounds produced by transformants with combinations of genes from the AspMel BGC or BeaBas BGC showed different optical properties with a gradually decreasing polarity over time. All samples, including the YMA control, have a small peak at retention time (RT) = 1.5 minutes, which is, due to its abundance, not included for further analysis.

HPLC traces of crude extracts from transformants with genes in the BeaBas BGC indicate presence of produced compounds. For all samples, except the YMA control, a single peak at RT = 2.1 min was detected (Fig. 15) (Table 4). In all samples, this detected peak consists of a compound that absorbs UV light with a wavelength of 215 and 268 nm (Fig. 16). Besides this peak, no strong signals of compounds were detected for the BeaBas pathway. Additionally, no mass of other compounds was detected using LC-MS (data not shown).

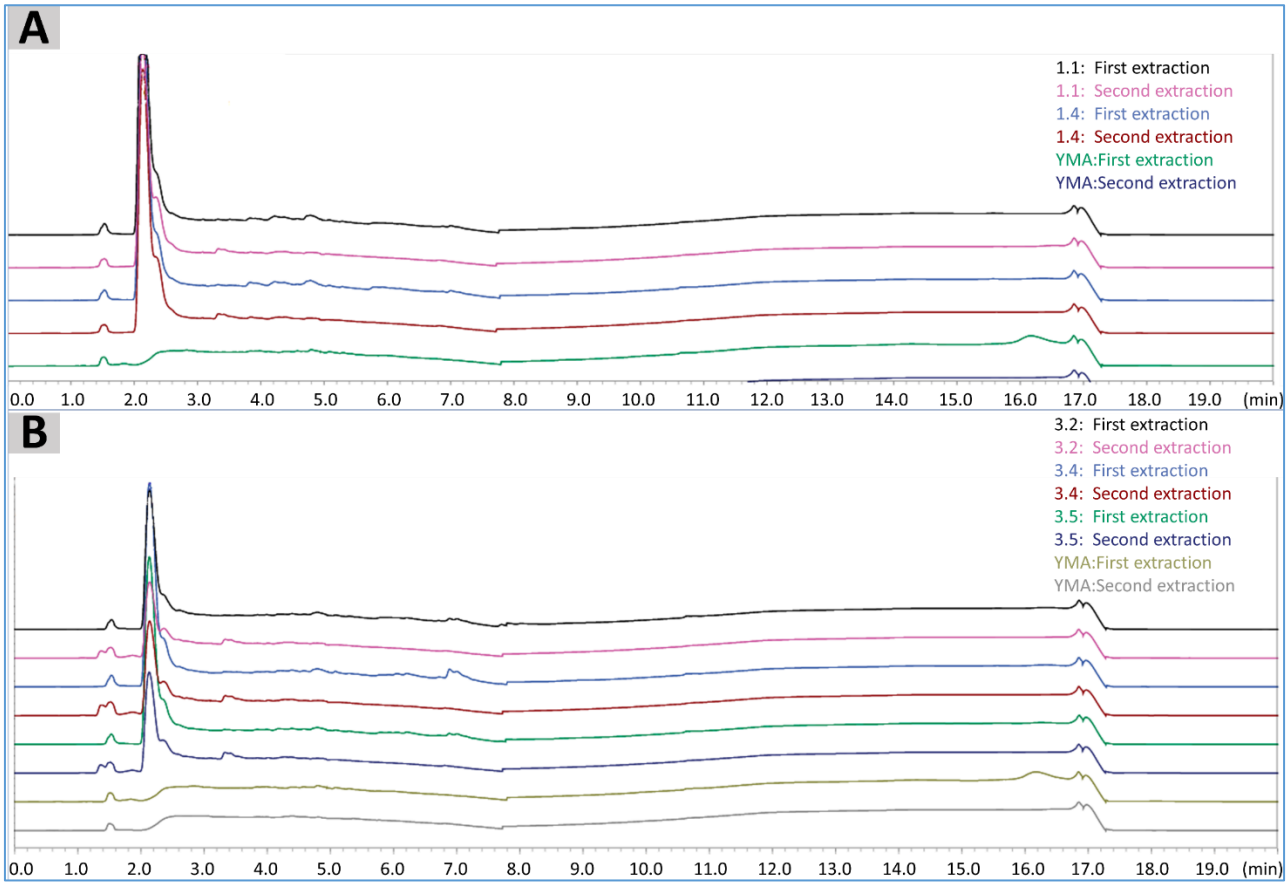


Figure 15. HPLC traces of crude extracts from *Aspergillus oryzae* NSAR1 transformants that confer genes of the BeaBas pathway. A) HPLC traces (190-800 nm) of *A. oryzae* transformed with pTYGSarg::BeaBas_*BBT1*, and of YMA; first and second round of metabolite extraction. B) HPLC traces (190-800 nm) of *A. oryzae* transformed with pTYGSade::BeaBas_*AMI*, and of YMA; first and second round of metabolite extraction. 1.1 & 1.4: pTYGSarg::BeaBas_*BBT1*; 3.2 & 3.4 & 3.5: pTYGSade::BeaBas_*AMI*; YMA: extraction from only YMA medium.

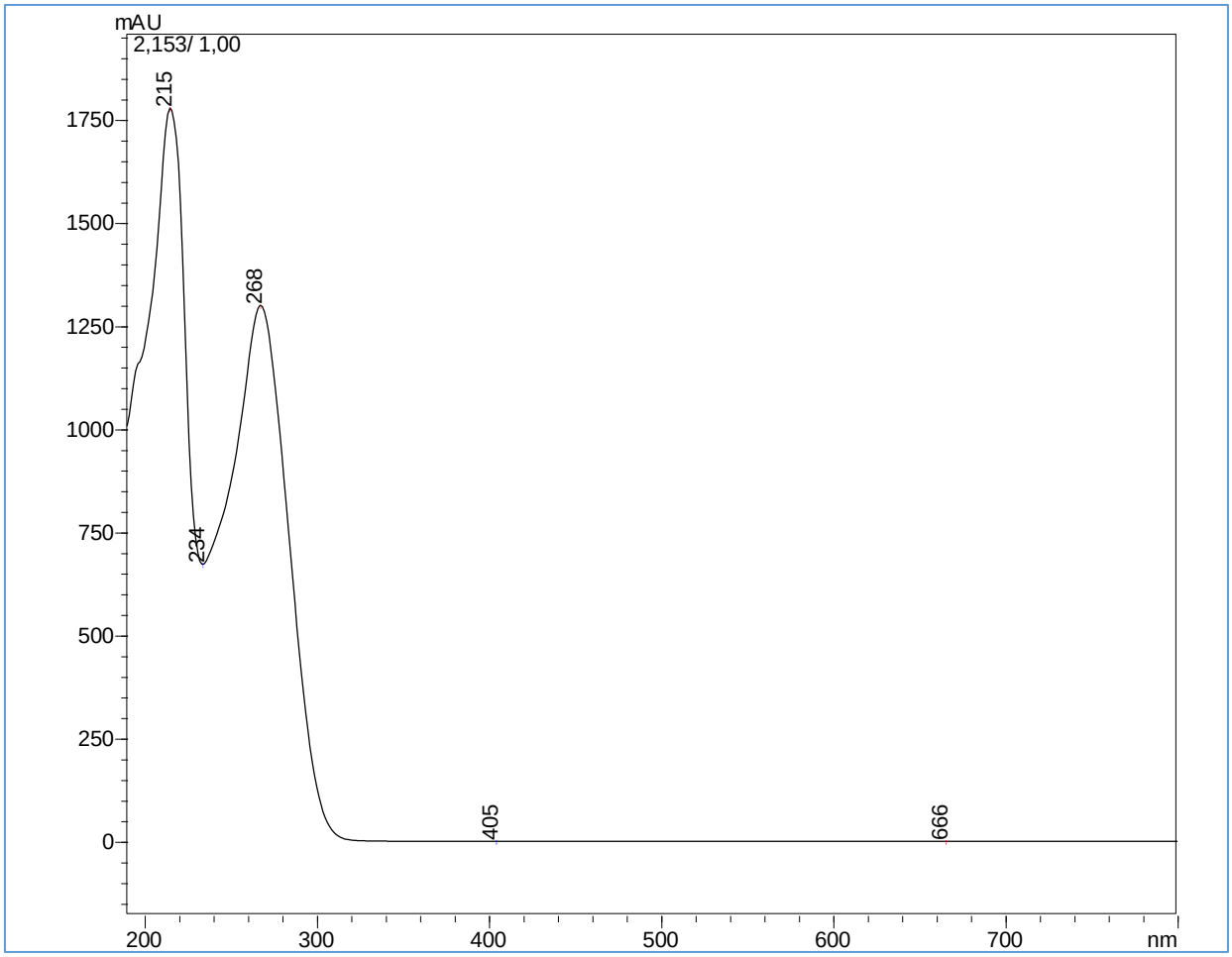


Figure 16. UV spectrum of sample 1.1* at retention time (RT) = 2.1 minutes. Compounds detected at RT = 2.1 minutes show absorbance at 215 and 268 nm. All crude extracts that have a peak at RT = 2.15 show the same UV spectrum. *Sample 1.1 was chosen for visualization

HPLC traces of crude extracts from transformants with genes of the AspMel pathway also showed peaks indicating that compounds were produced by the fungi. To start, the peak that was detected in transformants with genes from the BeaBas BGC at RT = 2.1 minutes with an absorption spectrum of 215 and 268 nm (Fig. 16**)** was also detected in several transformants with genes from the AspMel BGC (Table 4**)**.

Additionally, novel peaks were visible (Fig. 17). The results of the first and second extraction were not identical for all samples (Supplementary materials, Fig. S1), therefore only first extractions are compared for consistency. *A. oryzae* transformed with only tailoring genes, pTYGSade::AspMel_*ER* (8.3) and pTYGSarg::*F*:*M*:*Q* (10.1) did not show any new traces on HPLC or LC-MS (Fig. 17 and Supplementary materials, Fig. S2).

At RT = 4.6 minutes and RT = 5.4 minutes, a peak was detected for most samples transformed with the *PKS* from the AspMel BGC, *AME1* (Table 4). For *A. oryzae* NSAR1 transformed with only *AME1*, with or *AME1* and tailoring gene *ER* (6.1 and 4.1, respectively), the compounds detected at 4.6 minutes absorbed light at 201 and 275 nm (Fig. 18). The duplicates of transformations of *AME1* with *ER* (4.2) and solely *AME1* (5.2) did not show a peak besides the

one at RT = 2.1 minutes and were therefore excluded from further analysis (Supplementary materials, Fig. S1). Samples transformed with *AME1* and tailoring genes *FAD, MET,* and *QUEST* (5.1) showed a peak with wavelengths 200 and 286 nm (Fig. 19). The peak at 5.4 minutes shows compounds with wavelengths of 213/214, 260/261, and 298 nm if the fungi were only transformed with *AME1* or with both *AME1* and tailoring gene *ER*, respectively (Fig. 20). Of the latter, the LC-MS data showed that the compound at this retention time has a m/z of 167 detected on the negative mode (Fig. 21). The transformants with *AME1* and tailoring genes *FAD, MET,* and *QUEST* had a slightly different UV spectrum: the compound detected at 5.4 minutes absorbs light at wavelengths of 197, 212, 262, and 297 nm (Fig. 22). Additionally, it should be noted that the peak intensity of this sample at RT = 5.4 is relatively lower than the other present peaks compared to the samples transformed with only *AME1*, or with *AME1* and tailoring gene *ER* (Fig. 17). Unfortunately, no additional LC-MS data is available to compare the mass of the compounds at this RT between the transformants. Furthermore, the transformant with *AME1* and tailoring genes *FAD, MET,* and *QUEST* showed three unique peaks at RT = 4.0, RT = 5.2, and RT = 5.9 (Fig. 17). These peaks contain compounds absorbing light at 203 and 274 nm (Fig. 23A), 201 and 266 nm (Fig. 23B), and 201 and 270 nm (Fig. 23C), respectively.

Table 4. Overview of HPLC traces and LC-MS data of crude extracts from *Aspergillus oryzae* NSAR1 transformants that confer genes of the AspMel or BeaBas pathway. *A. oryzae* transformed with: 1.1 & 1.4: pTYGSarg::BeaBas_*BBT1*; 3.2 & 3.4 & 3.5: pTYGSade::BeaBas_*AMI*; 4.1 & 4.2: pTYGSarg::AspMel_*AME1* and pTYGSade::AspMel_*ER*; 5.1 & 5.2: pTYGSarg::AspMel_*AME1*::*F*:*M*:*Q*; 6.1: pTYGSarg::AspMel_*AME1*, 8.3 & 8.8 & 8.10: pTYGSade::AspMel_*ER*; 10.1: pTYGSarg::*F*:*M*:*Q*; first round of metabolite extraction.

Retention time (min)	Absorbance spectrum (nm)	Mass (m/z)	Sample
2.1	215, 268	-	1.1, 1.4 3.2, 3.4, 3.5 4.1, 4.2 5.2 8.3, 8.8, 8.10
4.0	203, 274	-	5.1
4.6	200, 286	-	5.1
4.6	201, 275	-	4.1 6.1
5.2	201, 266	-	5.1
5.4	214, 260, 298	167	4.1
5.4	213, 261, 298	-	6.1
5.4	197, 212, 262, 297	-	5.1
5.9	201, 270	-	5.1

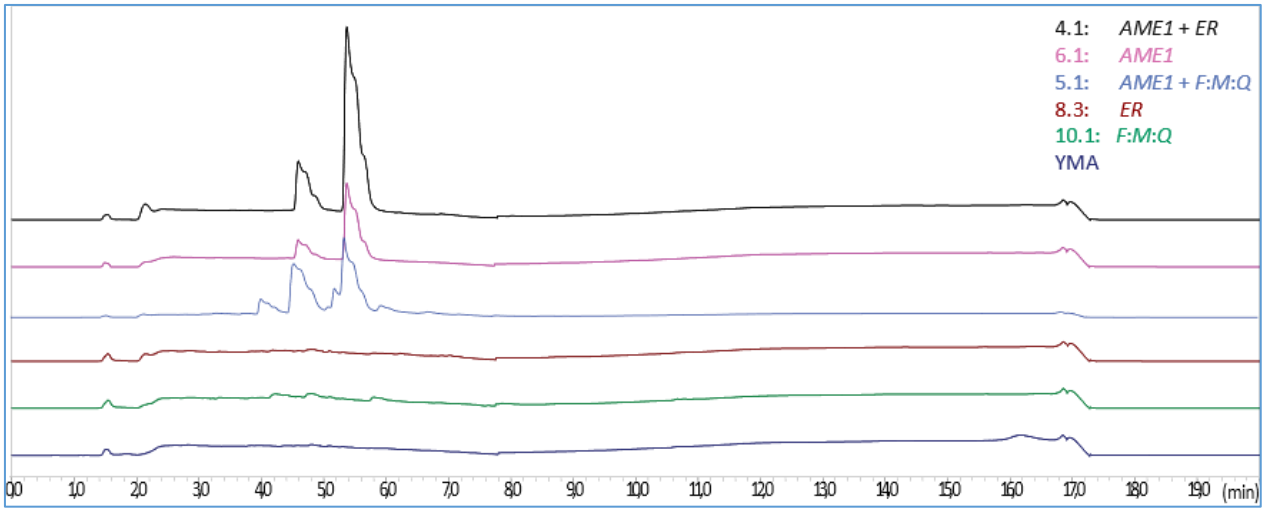


Figure 17. HPLC traces of crude extracts from *Aspergillus oryzae* NSAR1 transformants that confer genes of the AspMel pathway. HPLC traces (190-800 nm) of *A. oryzae* transformed with pTYGSarg::AspMel_*AME1* and pTYGSade::AspMel_*ER* (4.1), pTYGSarg::AspMel_*AME1* (6.1,) pTYGSarg::AspMel_*AME1*::*F*:*M*:*Q* (5.1), pTYGSade::AspMel_*ER* (8.3), pTYGSarg::*F*:*M*:*Q* (10.1), and of YMA; first round of metabolite extraction.

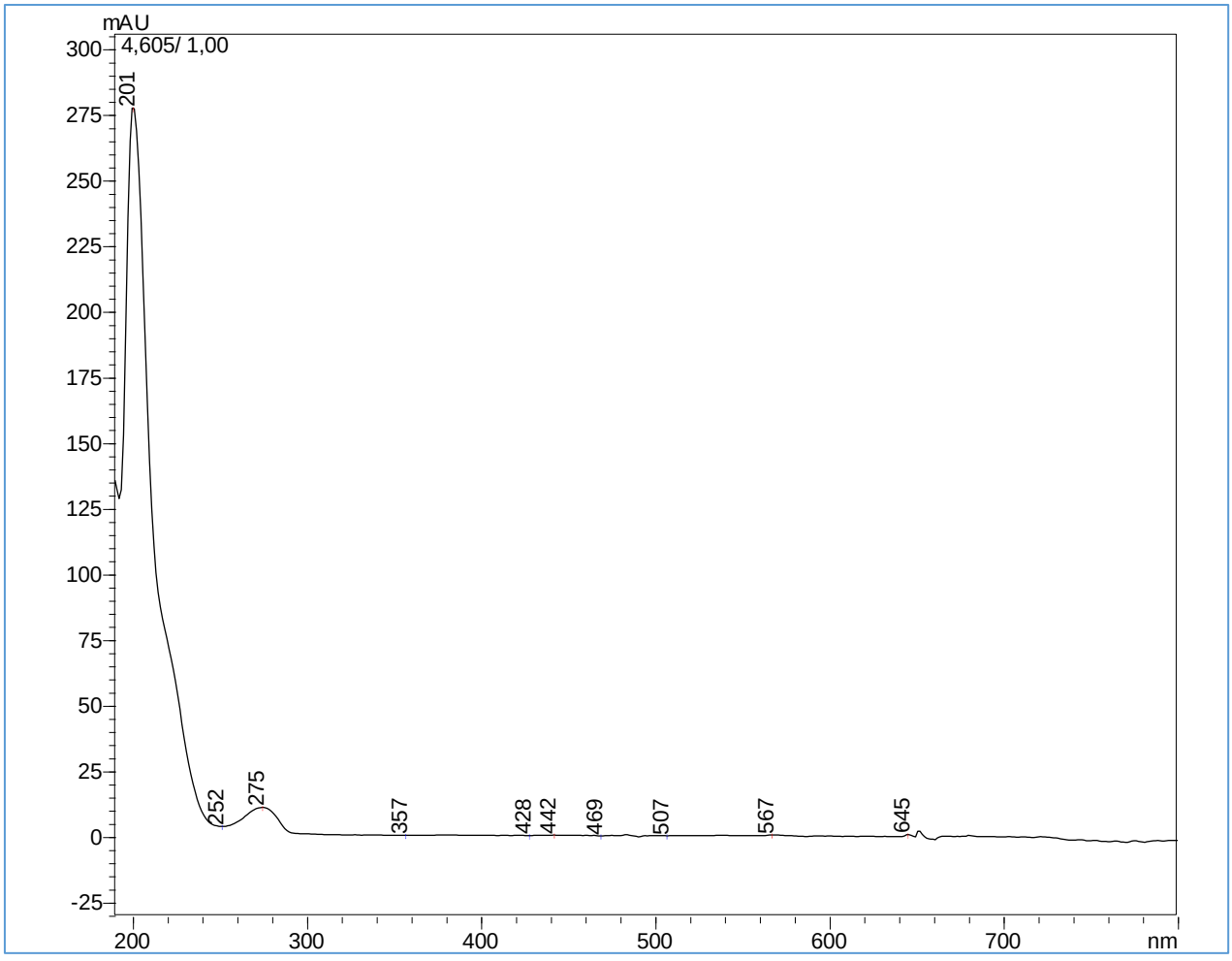


Figure 18. Absorbance spectrum of crude extracts from *Aspergillus oryzae* NSAR1 transformed with pTYGSarg::AspMel_*AME1* (6.1) at retention time = 4.6 minutes. NB this spectrum is the same for *A. oryzae* transformed with pTYGSarg::AspMel_*AME1* and pTYGSade::AspMel_*ER* (4.1) (data not shown).

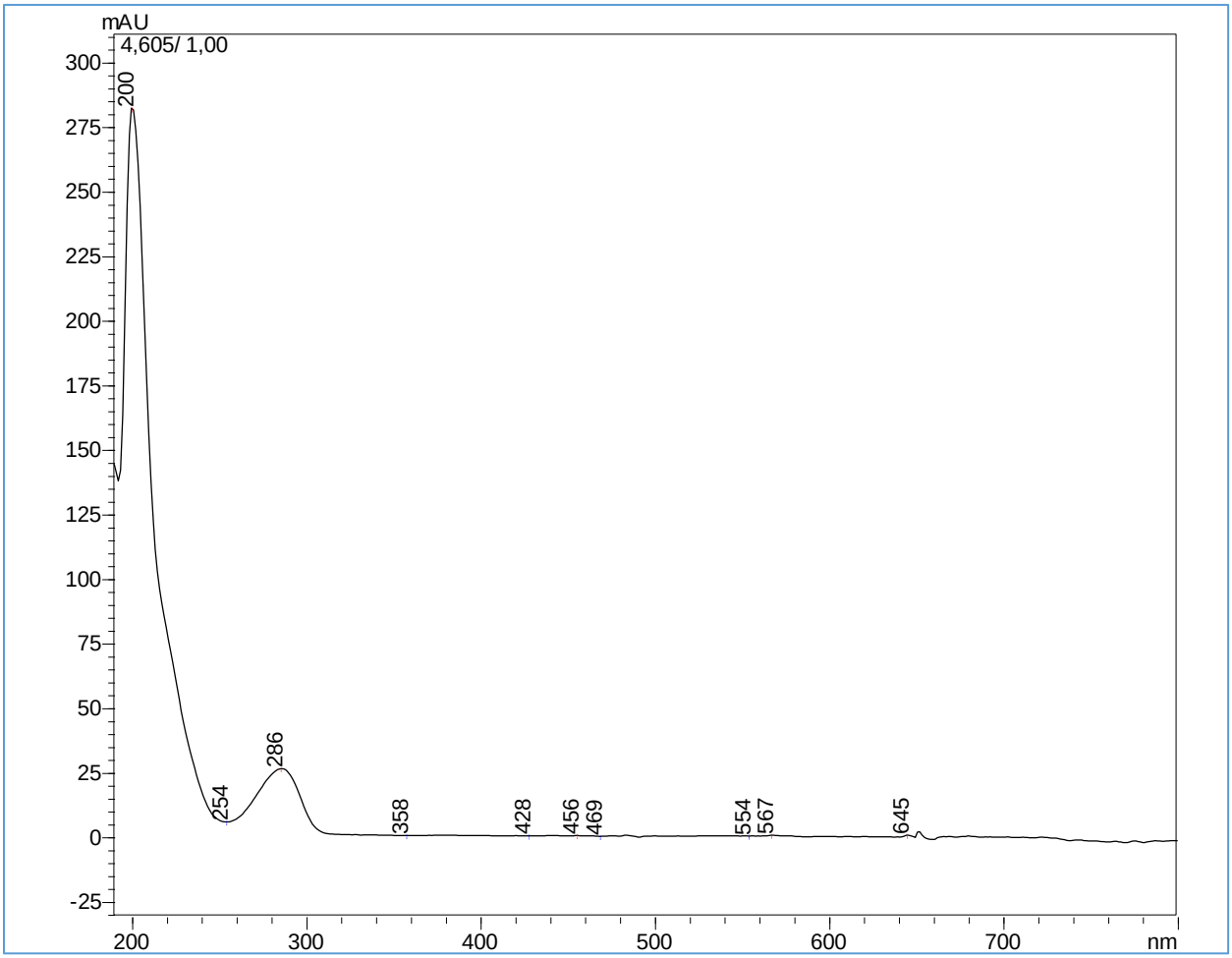


Figure 19. Absorbance spectrum of crude extracts from *Aspergillus oryzae* NSAR1 transformed with pTYGSarg::AspMel_*AME1*::*F*:*M*:*Q* (5.1) at retention time = 4.6 minutes.

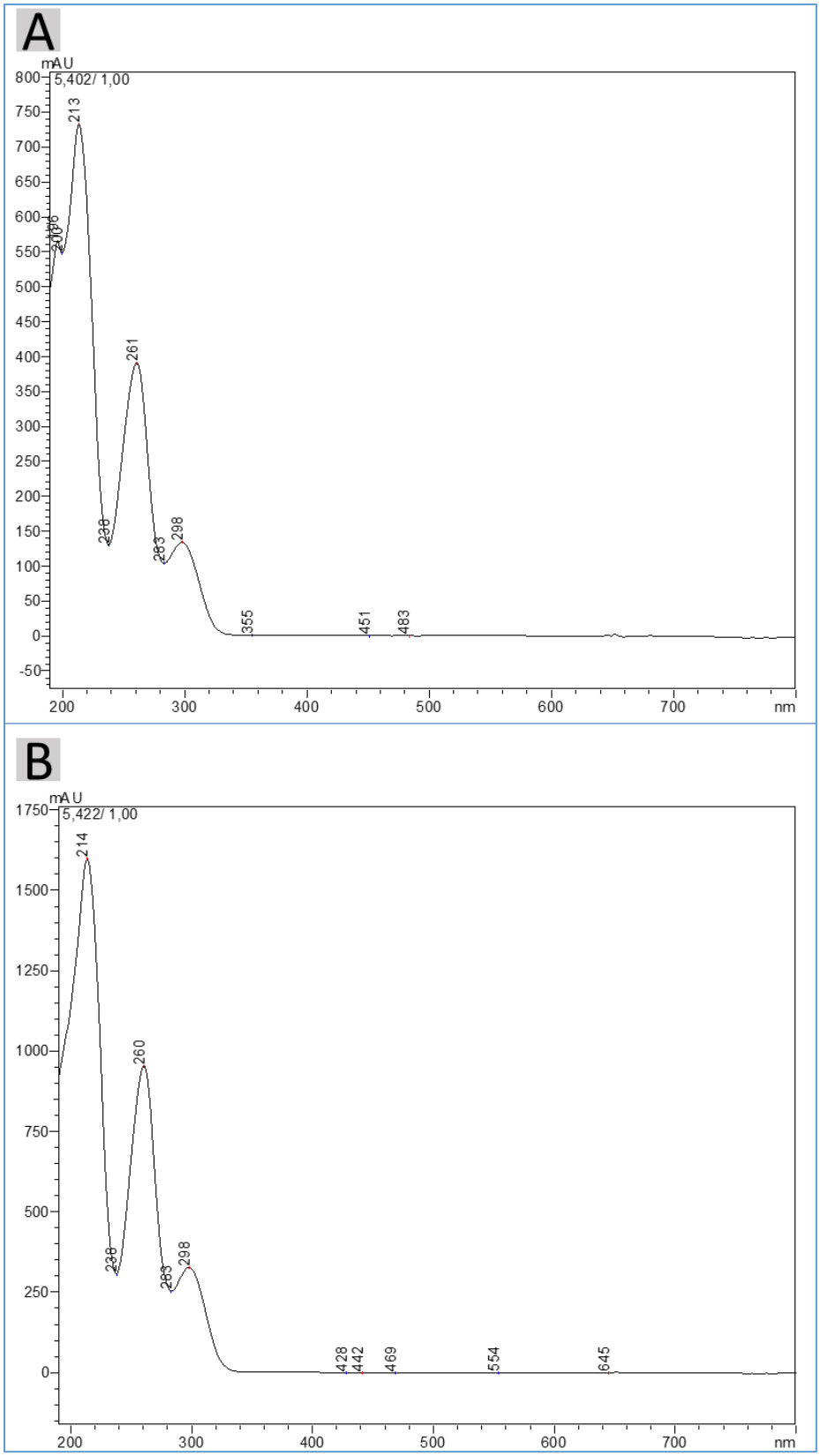


Figure 20. Absorbance spectrum of crude extracts from *Aspergillus oryzae* NSAR1 transformed with A) pTYGSarg::AspMel_*AME1* (6.1) and B) pTYGSarg::AspMel_*AME1* and pTYGSade::AspMel_*ER* (4.1) at retention time = 5.4 minutes.

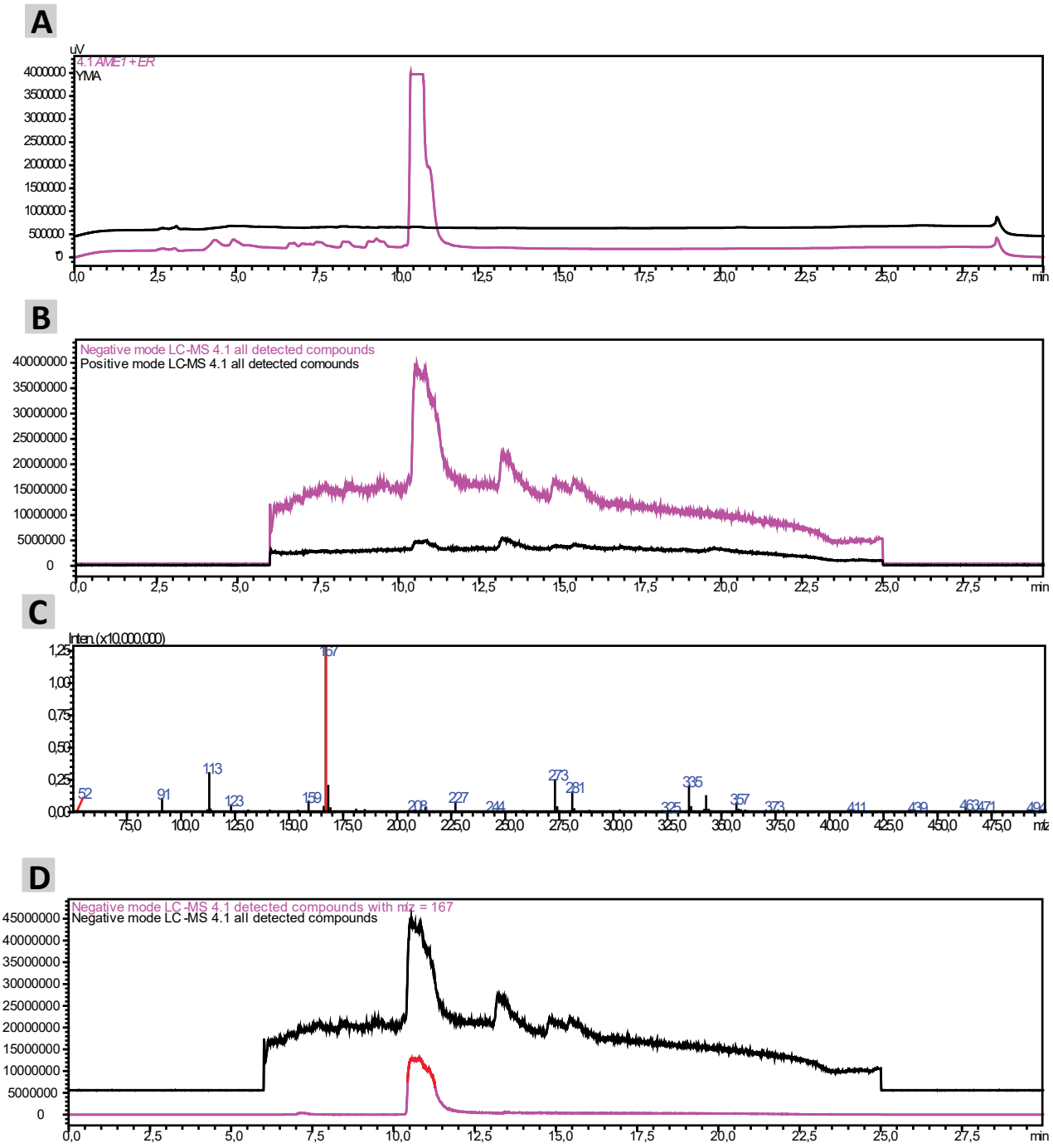


Figure 21. LC-MS data of crude extracts from *Aspergillus oryzae* NSAR1 transformed with pTYGSarg::AspMel_*AME1* and pTYGSade::AspMel_*ER* (4.1) linked to HPLC traces. A) HPLC traces (190-800 nm) of the crude extracts and of YMA; first round of extraction. B) LC-MS data of the crude extracts, on both positive and negative mode. C) Mass of LC-MS detected compounds at RT = 10.8 minutes. D) Presence of compounds with only an m/z of 167 in LC-MS data on the negative mode next to all detected compounds on the negative mode.

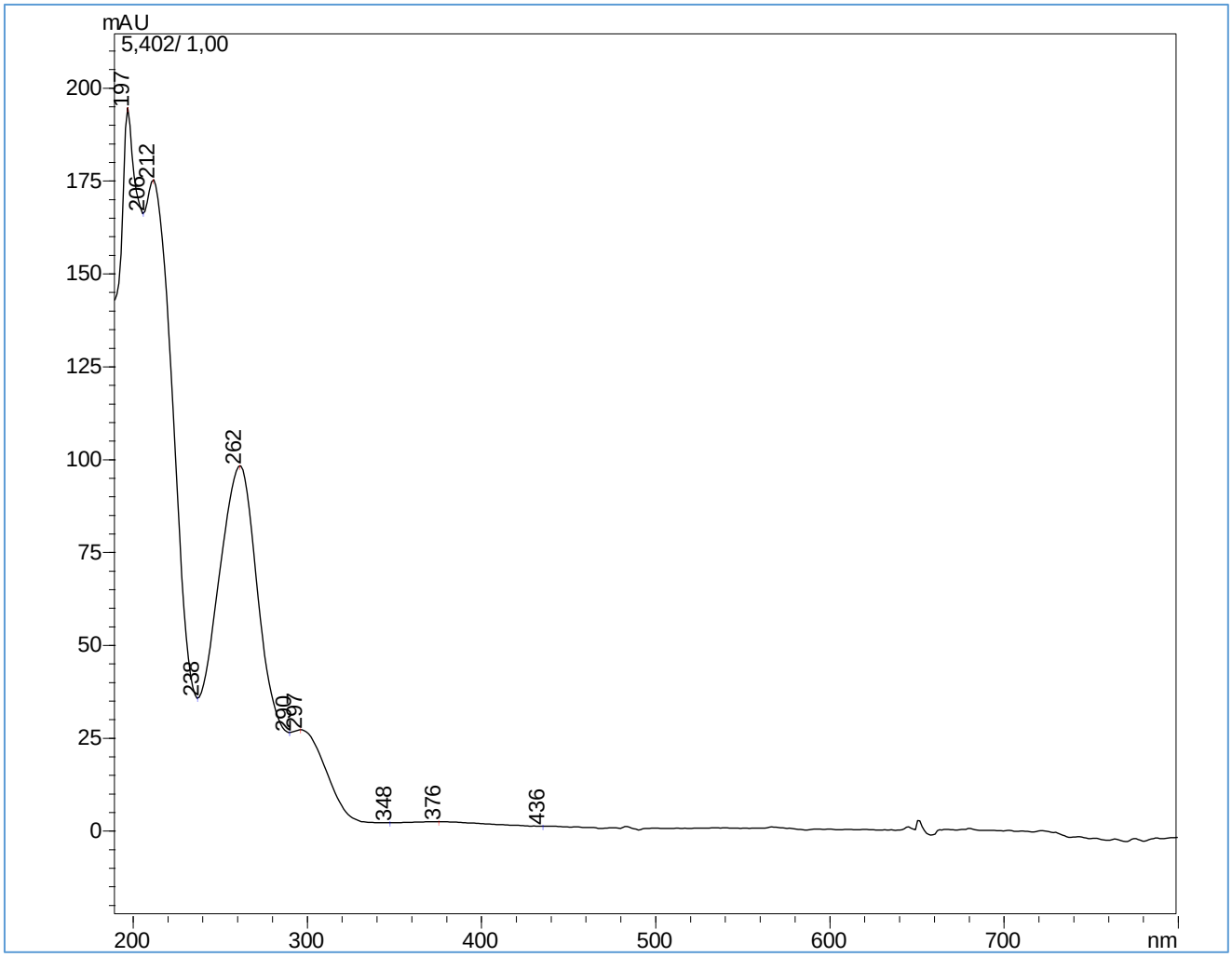


Figure 22. Absorbance spectrum of crude extracts from *Aspergillus oryzae* NSAR1 transformed with pTYGSarg::AspMel_*AME1*::*F*:*M*:*Q* (5.1) at retention time = 5.4 minutes.

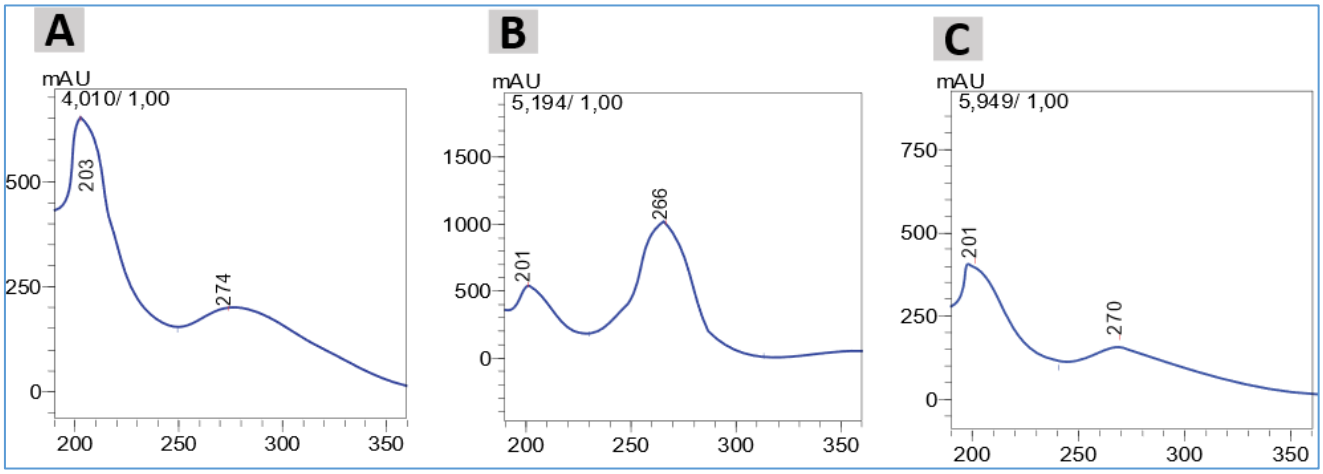


Figure 23. Absorbance spectrum of crude extracts from *Aspergillus* oryzae NSAR1 transformed with pTYGSarg::AspMel_*AME1*::*F*:*M*:*Q* (5.1) at retention time = 4.0 minutes (A), 5.2 minutes (B), and 5.9 minutes (C).

Discussion, conclusion, and recommendations

In this study, development of a tool to automate gene model curation was launched. Additionally, two biosynthetic gene clusters (AspMel and BeaBas) were expressed in *A. oryzae* NSAR1 to study the metabolic products of these BGCs.

Automated gene curation tool

The automated gene curation tool reads GenBank files given by the user as input data and stores the information. This information is used to select homologous gene models and the tool uses these gene models as reference of each other to evaluate the gene models' start, intron, and stop positions, by looking at the nucleotides in EBs. This developed tool is a set-up for gene-by-gene, alignment-based automated gene curation based on the hypothesis that errors in the predicted homologous gene models do not occur consistently. An odd number and minimum of 3 homologous gene models is needed to correct a gene model. Further development of the tool is required to accomplish this.

Evaluation of the EBs must be incorporated in the tool. This evaluation should check whether the sequences in the EBs all have the same length. If all sequences stop and end at the same position in the EB, this means that the prediction of this exon in the gene model is likely correct. If one or several, but not more than half of the sequences have a different length, the exon block is likely wrongly predicted and there can be several options which require different approaches for correction. For example, the first exon and the last exon require different checks as they have specific signature sequences. Also, it should be checked whether the different sequence is longer or shorter on the 5' end or the 3' end of the exon. Depending on this outcome, the exon block adjacent to this side must be checked, to see if for example an intron was missed in this sequence. Additionally, intron blocks (IBs) must be made using the same technique as the creation of EBs. These IBs can be used to re-align locally if needed to fix exon blocks. Additionally, they can serve as reference to check for intron signatures, such as the GT/GC at the donor site, AG at the acceptor site, and the CURAY branchpoint. The analysis results, changes, and corrected gene models can be the output of the tool when the development is finished.

Aspergillus oryzae NSAR1 transformation

A. oryzae NSAR1 was transformed with vectors containing genes from the chosen BGCs (AspMel and BeaBas), to afford heterologous expression of the BGCs and to be able to study the produced compounds. The transformation efficiency differed between the vectors and combinations of vectors. All transformations with only one vector produced colonies on the transformation plate, but not all colonies were able to grow after the second transfer to selection medium. These colonies might have been false positives on the transformation plate or may have been abortive transformants, which has been reported before [58]. Additionally, it is remarkable that the transformants with vectors containing a backbone of pTYGSade appeared to be more often true positives and mitotically more stable than the ones originating from pTYGSarg. Whereas all transformations with one vector had colonies grown on the transformation plate, co-transformations were less successful in general: of all co-transformations, only two colonies grew for the *A. oryzae* transformation with pTYGSarg::AspMel_*AME1* and pTYGSade::AspMel_*ER* (Table 3). Possibly, efficiency drops when performing transformations with multiple, larger

vectors. Previously, it was reported that 5 – 10 transformants per µg plasmid DNA of 5,500 bp were grown [58]. However, the vectors used in this study were much larger **(**Table 5**)** and during co-transformation, both vectors should be incorporated into the *A. oryzae* genome to be able to grow on the selection medium. pTYGSarg::AspMel_*AME1* and pTYGSade::AspMel_*ER* are relatively small plasmids in this study, which might be why this transformation was successful for two colonies. To overcome this inefficiency of co-transformations, existing transformed *A. oryzae* can be transformed again with another vector to gain a transformant that has been transformed with two vectors. This way, combinations of different vectors can be examined in future experiments. Furthermore, transformations with less than 3 transformants should be done again, to gain results in biological triplicate.

Table 5. Size of all constructed vectors used for *Aspergillus oryzae* NSAR1 in this study.

Plasmid name	Size (length in base pairs * 1000)
pTYGSarg::AspMel_*AME1*	21.1
pTYGSarg::AspMel_*AME1*::*F*:*M*:*Q*	23.9
PTYGSarg::*F*:*M*:*Q*	19.2
pTYGSade::AspMel_*ER*	14.3
pTYGSarg::BeaBas_*BBT1*	22.6
pTYGSade::BeaBas_*AMI*	18.1

After transformation, the produced compounds of the transformants were characterized. In the antimycotic disk diffusion assays against *C. albicans* and *P. rubens*, results were not consistent between crude extracts produced by *A. oryzae* NSAR1 transformed with genes from AspMel or BeaBas from the first and second round of metabolite extraction. These assays should therefore be repeated in triplicate. Additionally, the concentration amphotericin B should be re-evaluated. Since the positive control did not give consistent growth inhibition for the fungi, in this study 100 µL and 200 µL of two concentrations in the commonly used range (0.25 µg/mL and 2.5 µg/mL) were used on disk diffusion assays. These disks did not show any growth inhibition, which may be due to the concentration being too low. In other studies, 10 µg of amphotericin B is used during disk diffusion assays [59], whereas the in this study 0.025-0.5 µg was tested on the disk diffusion assays. Therefore, tests using a higher amount of amphotericin B will give more information on the concentration that is toxic for the fungal strains used at the Westerdijk Institute and will provide a reliable positive control.

Additionally, as mentioned before, the first and second round of metabolite extraction yielded not always consistent results in HPLC traces. Differences between the first and second metabolite extraction using EtAc can have several explanations. The compounds that migrate very easily to EtAc might be only present in the first extraction, while the ones that are more polar might be present in both extractions. The acidification before the second extraction can also affect the chemical properties of the compounds so they migrate more easily into EtAc during the second extraction, perhaps also changing HPLC detection or biological activity. To keep samples

comparable, the following discussion only considers crude extracts from the first round of metabolite extraction.

Looking at the HPLC traces, most of the transformants showed a peak at RT = 2.1 minutes with a UV spectrum of 214 and 268 nm, some very faint while others more intense (Table 4). This RT = 2.1 peak is likely kojic acid (KA), a secondary metabolite that is produced in large amounts by *A. oryzae* and has a UV spectrum with absorption bands at 215 and 269 nm [60], [61]. Therefore, this product is likely an SM originating from *A. oryzae* NSAR1 itself. This must be checked by chemical analysis of compounds produced by untransformed *A. oryzae* NSAR1, and also by *A. oryzae* NSAR1 transformed with empty pTYGSarg, pTYGSade, and a combination of the two. Furthermore, all visible peaks have a shoulder on the right side of the peak instead of being a single, pointy peak. This irregularity is due to technical issues with the column inside the HPLC machine.

Additionally, it is still necessary to perform PCR for all samples to check whether the genes have been inserted. Furthermore, expression of the integrated genes should be analyzed, for example by performing a reverse transcriptase quantitative PCR. If the gene is indeed expressed, it would also be interesting to check if the mRNA is translated to the protein to be able to link the phenotypic results with certainty to genomic, expression, and proteomic data.

Aspergillus melleus [RETRACTED] cluster [RETRACTED]

The disk diffusion assay of extracts from transformants expressing genes of this BGC showed that only sample 5.1, corresponding to *A. oryzae* NSAR1 transformed with pTYGSarg::AspMel_*AME1*::*F*:*M*:*Q*, had antibiotic activity. Since all HPLC traces of this sample are unique, the antibacterial compound is likely detected in one of those peaks. This could be studied by separation of the different peaks and using the separated extracts in a disk diffusion assay. Sample 5.2, transformed with the same vector, did not show any biological activity and besides KA, no produced compounds were detected with HPLC. However, 5.2 did not grow a second time on selection medium and is therefore thought to be false positive and not further taken into account. Furthermore, transformants containing vectors with only tailoring genes also did not show any biological activity or produced compounds, besides the KA peak. As the tailoring genes do generally not produce a metabolite, but rather modify the backbone assembled by core enzymes, the absence of bioactivity and novel compounds is expected. However, it should be checked if the genes are inserted, transcribed and the mRNA translated to verify these hypotheses.

When looked at the HPLC traces of crude extracts produced by *A. oryzae* NSAR1 transformed with *AME1* (6.1) or *AME1* and *ER* (4.1, 4.2), there were many similarities and differences. As the detection of compounds of transformants with *AME1* and *ER* was not reproducible in 4.2, only 4.1 was taken into account in this discussion. For both 6.1 and 4.1, a peak at RT = 4.6 minutes was detected of which the absorbance spectrum was the same (201 nm, 275 nm). The second peak present in 4.1 and 6.1 is at RT = 5.4 minutes. Again, for 6.1 and 4.1, the absorbance spectra were very similar (213/214, 261/260, and 298 nm for 6.1/4.1, respectively). From these results, it should be noted that the compounds produced by transformants with only *AME1* and transformants with both *AME1* and *ER* are not different in this analysis. Er can reduce an enoyl-thioester to an acyl moiety and in doing so it is expected that the UV spectrum would

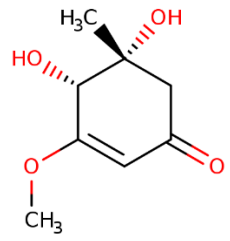
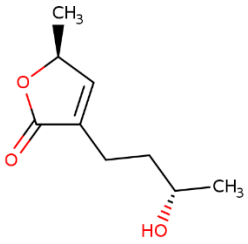
change [62]. However, this is not the case in 4.1. This could have several reasons, for example that *ER* is not correctly integrated or expressed by the fungus, which could be checked as described before. Secondly, it is possible that Er is expressed but did not cause any modifications in the compound, for example because an enoyl group is not present on the SM backbone, or the enoyl group is difficult to access for Er. Possibly Er is therefore not the first enzyme to modify the backbone in this biosynthetic pathway, or perhaps it is not involved at all. This can be tested by studying the compounds produced by transformed *A. oryzae* NSAR1 with different combinations of tailoring genes combined with *AME1*.

For sample 4.1, the compound detected by LC-MS at RT = 10.8 minutes is the same compound as the compound detected by HPLC at RT = 5.4 minutes. The m/z ratio of this product was 167 in the negative mode, meaning that the mass is 168 Da. This product is possibly the unmodified SM backbone of the AspMel BGC. The Natural Products Atlas [63] shows there is no compound known of this size originating from *A. melleus*. However, there is one aromatic compound known with mass 168 Da produced by a species within the *Aspergillus* genus (*A. giganteus*): Aspergilsmin C ($C_8H_8O_4$). However, a blast on the sequence of *AME1* using NCBI BLAST® [64] shows only one good alignment (79% percent identity): *A. steynii* [RETRACTED], annotated as putative polyketide synthase. A blast on the sequence of *AME1* using the JGI mycocosm [65] in the class of Eurotiomycetes gave more hits, of which 23 hits (including *A. neoauricomus* [RETRACTED], which is called *A. melleus* in this report and also has code [RETRACTED]) had a score of 1,569 or higher (Fig. 24). Although there are many hits, *A. giganteus* is not amongst them, hence it is not likely the 168 Da compound found in this study is the same as the already identified one originating from *A. giganteus*.

There are, however, 15 compounds found in a range of 158-178 Da in the family of *Aspergillus*, of which 2 with an aromatic compound (Table 6). It is however unknown what the absorption spectrum of these compounds is, and for (4S,5R)-4,5-Dihydroxy-3-methoxy-5-methylcyclohex-2-enone, it is unknown which species synthesize these metabolites. It is therefore likely that the backbone of the final SM is not yet linked to an *Aspergillus* species. However, tailoring enzymes can change the molecule and in doing so alter the mass, for example by adding a methyl group, which is likely the function of AspMel tailoring enzyme Met due to its O-methyl transferase domain. Additionally, as AspMel is the orthologous cluster of *R. peruviana* [RETRACTED], it would be interesting to search a larger database in future studies. Searching through more fungal species, for example in the phylum of ascomycetes, for alignments with *AME1* and linked SMs can give more insight in the discovery of compounds produced by orthologues of AspMel.

The preliminary results in this study indicate that the SM produced by genes in cluster [RETRACTED]of *A. melleus* might be a yet unidentified compound with antibacterial activity against gram negative (*E. coli*) and gram positive (*B. subtilis*) bacteria. Further experiments are needed to verify this.

Table 6. Aromatic compounds in the range of 158-178 Da produced by *Aspergillus* species.

Chemical structure and formula	Name (NPAID)	Molecular weight (Da)	Origin species	UV spectrum
 C8H12O4	(4S,5R)-4,5-Dihydroxy-3-methoxy-5-methylcyclohex-2-enone	172.18	*Aspergillus sp.*	Unknown
 C9H14O3	Aspilactonol A	170.208	*Aspergillus sp. 16-02-1*	Unknown

[FIGURE RETRACTED]

Figure 24. Blast results from JGI blastn with *AME1*. Only the hits with a score ≥1,569 are included in this figure. For more data, see Supplementary materials, Table S4.

Beauveria bassiana [RETRACTED] cluster [RETRACTED]

Disk diffusion assays against *E. coli* and *B. subtilis* showed that compounds produced by *A. oryzae* transformed with genes from BeaBas do not have any toxic effects against *E. coli* and *B. subtilis*. In the antimycotic assay, compounds from the first extraction produced by *A. oryzae* NSAR1 transformed with *BBT1* or *AMI* both showed growth inhibition on *C. albicans*. This is interesting, as the tailoring enzyme is not thought to produce a compound, but rather to modify the polyketide produced by Bbt1. To ensure the toxicity is caused by these tailoring genes, it would be necessary to transform *A. oryzae* with empty pTYGSade and pTYGSarg to study the possibility that *A. oryzae* itself produces toxic compounds to *C. albicans*. As an addition to this, it could be studied how Ami could modify the produced SMs of *A. oryzae* and check the bioactive and chemical properties of these resulting compounds. Additionally, as the co-transformation of *A. oryzae* NSAR1 with pTYGSarg::BeaBas_*BBT1* and pTYGSade::BeaBas_*AMI* was not successful in this study, another transformation round and characterization of the produced metabolites would be essential.

Furthermore, there is no clear unique peak in the crude extracts that might indicate expression of the BeaBas genes which cause production of an SM. However, since *BBT1* is annotated to code for a reducing PKS, the compound might be very reduced and therefore difficult to detect using the diode array UV HPLC. Fumonisins, for example, are known for this: lacking a useful chromophore or fluorophore due to the highly reduced nature of the compounds,

a different chemical analysis is required for chemical analysis, such as a HPLC with fluorescence detector [66]–[68]. Therefore, additional chemical analysis might be crucial to detect SMs from the BeaBas pathway. Additionally, assays to test biological activity of the produced compounds against adult insects and insect larvae can be interesting for this BGC, as the fungus *B. bassiana* is a known entomopathogen. To gain knowledge on this BGC might therefore also give more insight in the entomopathogenicity and could potentially be of importance to be applied in the agricultural sector as biopesticide.

Conclusion

In this study, the development of an automated gene curation cool was launched. The tool reads GenBank files given by the user, filters homologous gene models, and divides the gene models in so-called exon blocks. Further development is needed to evaluate and correct the exon blocks to incorporate actual automated gene curation.

Furthermore, *Aspergillus oryzae* was transformed with genes originating from a biosynthetic gene cluster of *Aspergillus melleus* (cluster [RETRACTED]) (AspMel) and of *Beauveria bassiana* (cluster [RETRACTED]) (BeaBas). Transformants with genes from BeaBas did not show any consistent antimicrobial activity or unique traces on the HPLC or LC-MS. Transformants with the AspMel non-reducing polyketide synthase (*AME1*) and three tailoring genes (*FAD binding protein, O-methyltransferase*, and *questin oxidase-like*) showed antibiotic activity against *Escherichia coli* and *Bacillus subtilis*. Additionally, this sample showed four unique peaks detected by HPLC. Compounds were detected on HPLC for the transformants with *AME1* alone or *AME1* and *enoyl reductase*, of which the latter produces a compound of 168 Da that is likely not identified yet. Further experiments are needed to verify these results and to gain more knowledge on the biosynthetic pathways of both BGCs.

Bibliography

[1] H.-W. Nützmann, C. Scazzocchio, and A. Osbourn, "Metabolic gene clusters in eukaryotes," *Annu. Rev. Genet.*, vol. 52, pp. 159–183, 2018.

[2] A. Rokas, J. H. Wisecaver, and A. L. Lind, "The birth, evolution and death of metabolic gene clusters in fungi," *Nat. Rev. Microbiol.*, vol. 16, no. 12, pp. 731–744, 2018.

[3] A. L. Harvey, R. Edrada-Ebel, and R. J. Quinn, "The re-emergence of natural products for drug discovery in the genomics era," *Nat. Rev. Drug Discov.*, vol. 14, no. 2, pp. 111–129, 2015.

[4] M. P. Narsing Rao, M. Xiao, and W.-J. Li, "Fungal and bacterial pigments: Secondary metabolites with wide applications," *Front. Microbiol.*, vol. 8, Jun. 2017.

[5] K. D. Hyde *et al.*, "The amazing potential of fungi: 50 ways we can exploit fungi industrially," *Fungal Divers.*, vol. 97, no. 1, pp. 1–136, 2019.

[6] F. A. Nagia and R. S. R. El-Mohamedy, "Dyeing of wool with natural anthraquinone dyes

from Fusarium oxysporum," *Dyes and Pigments*, vol. 75, no. 3, pp. 550-555, 2006.

[7] M. E. Zain, "Impact of mycotoxins on humans and animals," *J. Saudi Chem. Soc.*, vol. 15, no. 2, pp. 129–144, 2011.

[8] C. De Costa, "St Anthony's fire and living ligatures: a short history of ergometrine," *Lancet*, vol. 359, pp. 1768–1770, 2002.

[9] A. Grzybowski, K. Pawlikowska-łagód, and A. Polak, "Ergotism and Saint Anthony's Fire," *Clin. Dermatol.*, 2021.

[10] A. -un-Nisa, N. Zahra, Y. Nazir Butt, and S. Hina, "Aflatoxins; a potential threat to human health," *Pakistan J. Food Sci.*, vol. 24, no. 4, pp. 256–271, 2014.

[11] J. L. Richard, "Discovery of aflatoxins and significant historical features," *Toxin Rev.*, vol. 27, pp. 171–201, Jan. 2008.

[12] N. P. Keller, "Fungal secondary metabolism: Regulation, function and drug discovery," *Nat. Rev. Microbiol. 2018 173*, vol. 17, no. 3, pp. 167–180, Dec. 2018.

[13] J. Houbraken, J. C. Frisvad, and R. A. Samson, "Fleming's penicillin producing strain is not Penicillium chrysogenum but P. rubens," *IMA Fungus*, vol. 2, no. 1, pp. 87–95, 2011.

[14] M. Stadler and P. Dersch, *How to Overcome the Antibiotic Crisis*. 2016.

[15] J. W-H Li and J. C. Vederas, "Drug discovery and natural products: End of an era or an endless frontier?," *Science*, vol. 325, pp. 161–165.

[16] A. Maxmen, "Crop pests: Under attack," *Nature*, vol. 501, pp. S15–S17, Sep. 2013.

[17] S. Kumar, "Biopesticides: A need for food and environmental safety," *J. Biofertilizers Biopestic.*, vol. 03, no. 04, 2012.

[18] A. Hussain, M. Rizwan-ul-Haq, H. Al-Ayedh, and A. Al-Jabr, "Mycoinsecticides: Potential and future perspective," *Recent Pat. Food. Nutr. Agric.*, vol. 6, no. 1, pp. 45–53, 2014.

[19] O. Mosunova, J. C. Navarro-Muñoz, and J. Collemare, "The biosynthesis of fungal secondary metabolites: From fundamentals to biotechnological applications," *Ref. Modul. Life Sci.*, pp. 1–19, 2020.

[20] N. P. Keller and T. M. Hohn, "Metabolic pathway gene clusters in filamentous fungi," *Fungal Genet. Biol.*, vol. 21, pp. 17–29, 1997.

[21] A. A. Brakhage and V. Schroeckh, "Fungal secondary metabolites-strategies to activate silent gene clusters," *Fungal Genet. Biol.*, vol. 48, pp. 15–22, 2011.

[22] D. Hoffmeister and N. P. Keller, "Natural products of filamentous fungi: Enzymes, genes, and their regulation," *Nat. Prod. Rep.*, vol. 24, no. 2, pp. 393–416, 2007.

[23] R. J. Cox, “Polyketides, proteins and genes in fungi: Programmed nano-machines begin to reveal their secrets,” *Org. Biomol. Chem.*, vol. 5, pp. 2010–2026, 2007.

[24] K. E. Bushley and G. Turgeon, “Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships,” *BMC Evol. Biol.*, vol. 10, 2010.

[25] K. M. Fisch, “Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS,” *RSC Adv.*, vol. 3, no. 40, pp. 18228–18247, Sep. 2013.

[26] D. A. Herbst, C. A. Townsend, and T. Maier, “The architectures of iterative type I PKS and FAS,” *Nat. Prod. Rep.*, vol. 35, no. 10, pp. 1046–1069, Oct. 2018.

[27] J. C. Navarro-Muñoz and J. Collemare, “Evolutionary histories of type III polyketide synthases in fungi,” *Front. Microbiol.*, vol. 10, no. January, pp. 1–18, 2020.

[28] Y. Shimizu, H. Ogata, and S. Goto, “Type III polyketide synthases: functional classification and phylogenomics,” *ChemBioChem*, vol. 18, pp. 50–65, 2017.

[29] N. Khaldi *et al.*, “SMURF: genomic mapping of fungal secondary metabolite clusters,” *Fungal Genet. Biol.*, vol. 47, no. 9, p. 736, Sep. 2010.

[30] K. Blin *et al.*, “antiSMASH 6.0: improving cluster detection and comparison capabilities,” *Nucleic Acids Res.*, vol. 49, 2021.

[31] I. Kjaerbølling, U. H. Mortensen, T. Vesth, and M. R. Andersen, “Strategies to establish the link between biosynthetic gene clusters and secondary metabolites,” *Fungal Genet. Biol.*, vol. 130, pp. 107–121, 2019.

[32] G. D. Hannigan *et al.*, “A deep learning genome-mining strategy for biosynthetic gene cluster prediction,” *Nucleic Acids Res.*, vol. 47, no. 18, pp. e110–e110, Oct. 2019.

[33] M. Adamek *et al.*, “Comparative genomics reveals phylogenetic distribution patterns of secondary metabolites in Amycolatopsis species,” *BMC Genomics*, vol. 19, pp. 1–15, Jun. 2018.

[34] R. L. Bertrand and J. L. Sorensen, “A comprehensive catalogue of polyketide synthase gene clusters in lichenizing fungi,” *J. Ind. Microbiol. Biotechnol.*, vol. 45, no. 3, pp. 1067–1081, 2080.

[35] M. Adamek, M. Alanjary, and N. Ziemert, “Applied evolution: phylogeny-based approaches in natural products research,” *Nat. Prod. Rep.*, vol. 36, pp. 1295–1312, 2019.

[36] L. Liu, Z. Zhang, C.-L. Shao, J.-L. Wang, H. Bai, and C.-Y. Wang, “Bioinformatical analysis of the sequences, structures and functions of fungal polyketide synthase product template domains,” *Sci. Reports 2015 51*, vol. 5, no. 1, pp. 1–12, May 2015.

[37] E. Skellam, “Strategies for engineering natural product biosynthesis in fungi,” *Trends*

Biotechnol., vol. 37, pp. 416–427, 2019.

[38] S. Bergmann, J. Schümann, K. Scherlach, C. Lange, A. A. Brakhage, and C. Hertweck, "Genomics-driven discovery of PKS-NRPS hybrid metabolites from Aspergillus nidulans," *Nat. Chem. Biol.*, vol. 3, no. 4, pp. 213–217, Mar. 2007.

[39] K. A. K. Pahirulzaman, K. Williams, and C. M. Lazarus, "A toolkit for heterologous expression of metabolic pathways in Aspergillus oryzae", *Methods in Enzymology*, vol. 517, pp. 241-260, 2012.

[40] F. Alberti, G. D. Foster, and A. M. Bailey, "Natural products from filamentous fungi and production by heterologous expression," *Appl. Microbiol. Biotechnol.*, vol. 101, no. 2, pp. 493–500, Dec. 2016.

[41] C. M. Lazarus, K. Williams, and A. M. Bailey, "Reconstructing fungal natural product biosynthetic pathways," *Nat. Prod. Rep.*, vol. 31, no. 10, pp. 1339–1347, 2014.

[42] A. van der Burgt, E. Severing, J. Collemare, and P. J. G. M. de Wit, "Automated alignment-based curation of gene models in filamentous fungi," *BMC Bioinformatics*, vol. 15, pp. 1–13, 2014.

[43] Y. F. Li *et al.*, "Comprehensive curation and analysis of fungal biosynthetic gene clusters of published natural products HHS Public Access," *Fungal Genet Biol*, vol. 89, pp. 18–28, 2016.

[44] R. Lücking, B. P. Hodkinson, and S. D. Leavitt, "The 2016 classification of lichenized fungi in the Ascomycota and Basidiomycota – Approaching one thousand genera," *Bryologist*, vol. 119, no. 4, pp. 361–416, Jan. 2017.

[45] C. A. Borrias, "Activating silent fungal biosynthetic gene clusters in search for novel antimicrobial compounds," *Utr. Univ. Intern. Doc.*, 2021.

[46] J. T. Kealey, J. P. Craig, and P. J. Barr, "Identification of a lichen depside polyketide synthase gene by heterologous expression in Saccharomyces cerevisiae," *Metab. Eng. Commun.*, vol. 13, p. e00172, Dec. 2021.

[47] W. Kim *et al.*, "Linking a gene cluster to atranorin, a major cortical substance of lichens, through genetic dereplication and heterologous expression," *MBio*, vol. 12, no. 3, Jun. 2021.

[48] R. L. Bertrand and J. L. Sorensen, "Lost in translation: challenges with heterologous expression of lichen polyketide synthases," *Chem. Sel.*, vol. 4, pp. 6473–6483, 2019.

[49] A. Amobonye, P. Bhagwat, A. Pandey, S. Singh, and S. Pillai, "Biotechnological potential of *Beauveria bassiana* as a source of novel biocatalysts and metabolites," *Critical reviews in Biotechnoloy*, vol. 40, pp. 1019-1034, 2020.

[50] S. Udompaisarn *et al.*, “The polyketide synthase PKS15 has a crucial role in cell wall formation in *Beauveria bassiana*,” *Sci. Rep.*, vol. 10, 2020.

[51] J. Punya *et al.*, “Phylogeny of type I polyketide synthases (PKSs) in fungal entomopathogens and expression analysis of PKS genes in *Beauveria bassiana* BCC 2660,” *Fungal Biol.*, vol. 119, pp. 538–550, 2015.

[52] Y. Zhang, U. Werling, and W. Edelmann, “SLiCE: a novel bacterial cell extract-based DNA cloning method,” *Nucleic Acids Res.*, vol. 40, no. 8, pp. e55–e55, Apr. 2012.

[53] D. K. Williams *et al.*, “Heterologous production of fungal maleidrides reveals the cryptic cyclization involved in their biosynthesis,” *Angew. Chem. Int. Ed. Engl.*, Jun. 2016.

[54] J. D. Tripp, J. L. Lilley, W. N. Wood, and L. K. Lewis, “Enhancement of plasmid DNA transformation efficiencies in early stationary-phase yeast cell cultures,” *Yeast*, vol. 30, no. 5, pp. 191–200, May 2013.

[55] J. Rozewicki, S. Li, K. M. Amada, D. M. Standley, and K. Katoh, “MAFFT-DASH: integrated protein sequence and structural alignment,” *Nucleic Acids Res.*, vol. 47, no. W1, pp. W5–W10, Jul. 2019.

[56] S. Capella-Gutiérrez, J. M. Silla-Martínez, and T. Gabaldón, “trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses,” *Bioinforma. Appl. NOTE*, vol. 25, no. 15, pp. 1972–1973, 2009.

[57] M. P. Dunne and S. Kelly, “OMGene: mutual improvement of gene models through optimisation of evolutionary conservation,” *BMC Genomics*, vol. 19, pp. 1–18, Apr. 2018.

[58] Y. T. Hahm and C. A. Batt, “ Genetic transformation of an argB mutant of *Aspergillus oryzae* ,” *Appl. Environ. Microbiol.*, vol. 54, no. 6, pp. 1610–1611, Jun. 1988.

[59] S. S. W. Wong *et al.*, “*In vitro* and *in vivo* activity of a novel antifungal small molecule against *Candida* infections,” *PLoS One*, vol. 9, no. 1, Jan. 2014.

[60] K. Tamano *et al.*, “Identification and characterization of genes responsible for biosynthesis of kojic acid, an industrially important compound from *Aspergillus oryzae*,” *Fungal Genet. Biol.*, 2010.

[61] M. S. Piantavini *et al.*, “Elucidation of the electronic spectrum changes of Ka-Al $^{3+}$ complex by potentiometric titration, FTIR, ^{13}C RMN and quantum mechanics,” *Quim. Nov.*, vol. 40, no. 7, pp. 774–780, 2017.

[62] R. P. Massengo-Tiassø and J. E. Cronan, “Diversity in enoyl-acyl carrier protein reductases,” *Cell. Mol. Life Sci.*, vol. 66, pp. 1507–1517, 2009.

[63] J. A. Van Santen *et al.*, “The Natural Products Atlas: An open access knowledge base for

microbial natural products discovery," *ACS Cent. Sci.*, vol. 5, pp. 1824–1833, 2019.

[64] "BLAST: Basic Local Alignment Search Tool." [Online]. Available: https://blast.ncbi.nlm.nih.gov/Blast.cgi. [Accessed: 28-Aug-2021].

[65] I. V. Grigoriev *et al.*, "MycoCosm portal: Gearing up for 1000 fungal genomes," *Nucleic Acids Res.*, vol. 42, no. D1, Jan. 2014.

[66] N. Ndube, L. Van Der Westhuizen, and G. S. Shephard, "Determination of fumonisins in maize by HPLC with ultraviolet detection of o-phthaldialdehyde derivatives," *Mycotoxin Res.*, vol. 25, no. 225–228, 2009.

[67] B. Gaspardo *et al.*, "A rapid method for detection of fumonisins B 1 and B 2 in corn meal using Fourier transform near infrared (FT-NIR) spectroscopy implemented with integrating sphere," *Food Chem.*, vol. 135, pp. 1608–1612, 2012.

[68] X. Zhu, C. Vogeler, and L. Du, "Functional complementation of Fumonisin biosynthesis in *FUM1*-disrupted *Fusarium verticillioides* by the AAL-toxin polyketide synthase gene *ALT1* from *Alternaria alternata* f. sp. *Lycopersici*," *J. Nat. Prod.*, vol. 71, pp. 957–960, 2008.

Supplementary materials

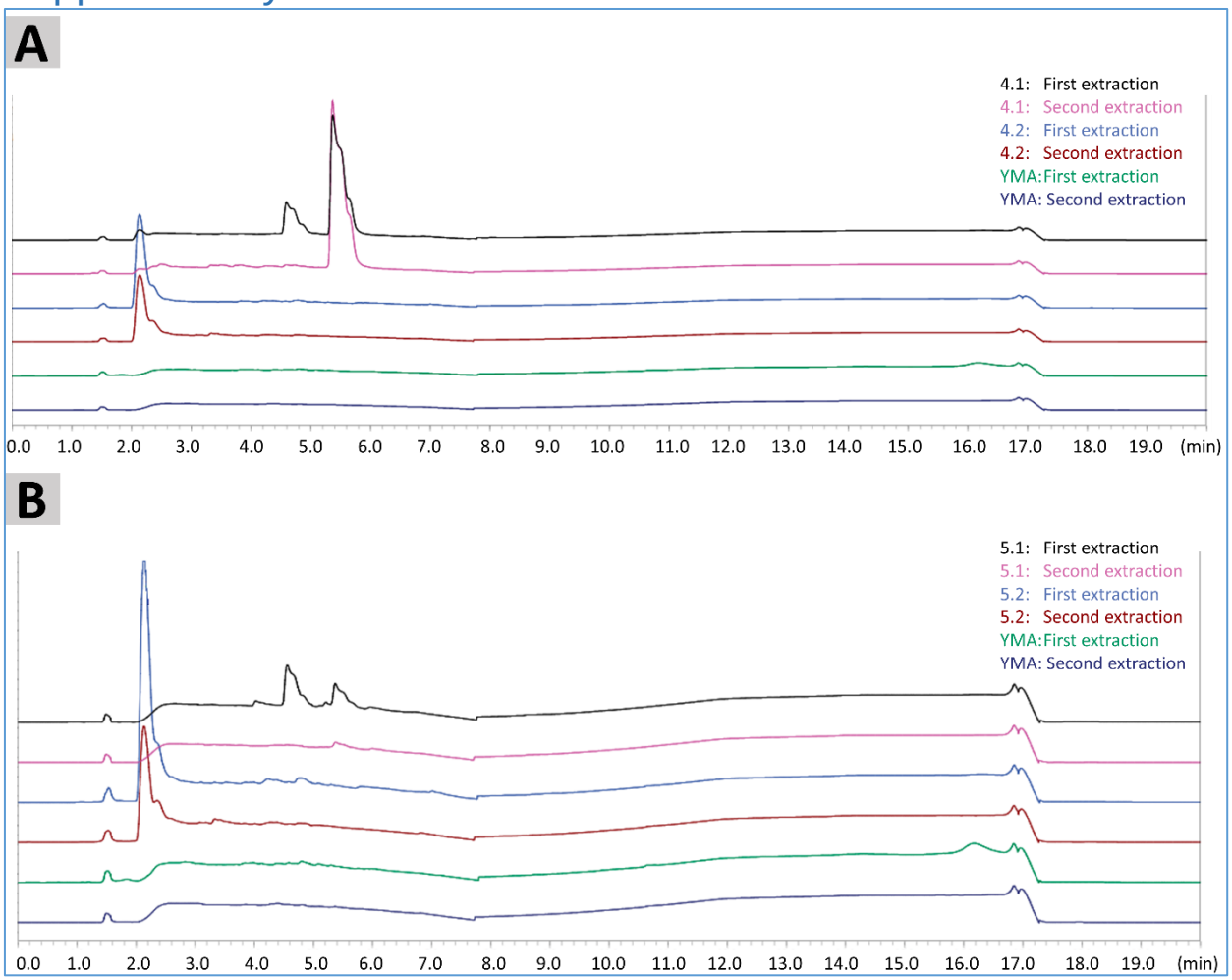


Figure S1. HPLC traces of crude extracts from Aspergillus oryzae NSAR1 transformants that confer genes of the AspMel pathway. A) HPLC traces (190-800 nm) of A. oryzae transformed with pTYGSarg::AspMel_AME1 and pTYGSade::AspMel_ER (4.1 and 4.2) and of YMA; first and second round of extraction. B) HPLC traces (190-800 nm) of A. oryzae transformed with pTYGSarg::AspMel_AME1::F:M:Q (5.1 and 5.2) and of YMA; first round of metabolite extraction.

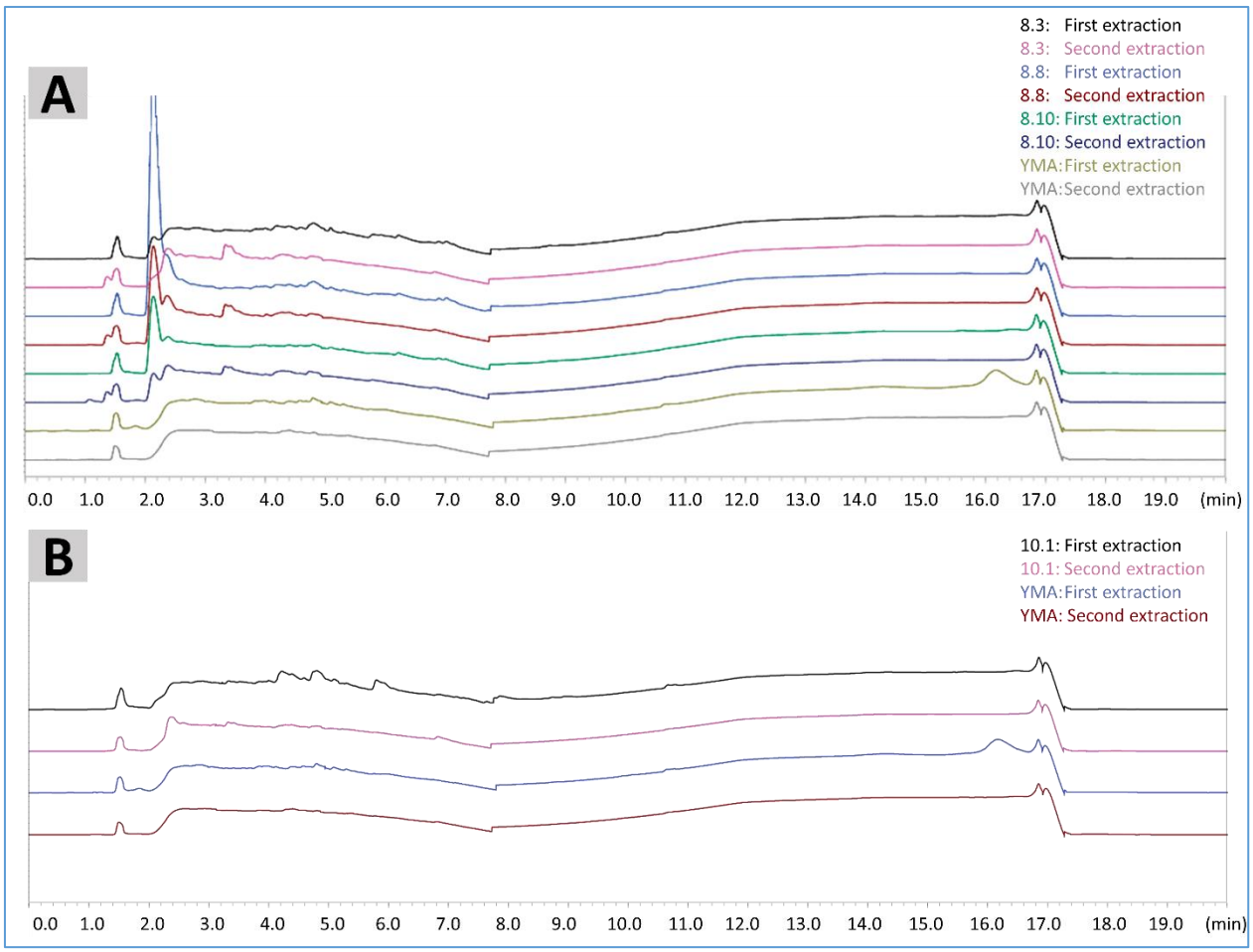


Figure S2. HPLC traces of crude extracts from *Aspergillus* oryzae NSAR1 transformants that confer genes of the AspMel pathway. A) HPLC traces (190-800 nm) of *A. oryzae* transformed with pTYGSade::AspMel_*ER* and of YMA; first and second round of extraction. B) HPLC traces (190-800 nm) of *A. oryzae* transformed with pTYGSade::AspMel_*ER* and of YMA; first and second round of extraction pTYGSarg::*F*:*M*:*Q* (10.1), and of YMA; first round of metabolite extraction.

Table S1. Overview of conserved domain of the genes within the chosen biosynthetic gene clusters.

Organism	Cluster	Gene ID	Predicted domains
Aspergillus melleus **[RETRACTED]**	[RETRACTED]	[RETRACTED]	[RETRACTED]
		[RETRACTED]	[RETRACTED]
		[RETRACTED]	[RETRACTED]
		[RETRACTED]	[RETRACTED]
		[RETRACTED]	[RETRACTED]
Beauveria bassiana **[RETRACTED]**	[RETRACTED]	[RETRACTED]	[RETRACTED]
		[RETRACTED]	[RETRACTED]

Table S2. Recipes for media used in this study.

Medium	Components (Manufacturer and product code)	Amount
Lysogeny agar (LA)	Tryptone (Oxoid LP0042)	10.0 gr
	Yeast extract (Difco 212750)	5.0 gr
	Sodium chloride (NaCl) (Baker 0278)	5.0 gr
	Thymine (Sigma T-0376)	0.02 gr
	Agar bacteriological, Agar No.1 (Oxoid L11)	15.0 gr
	Demi H_2O	1.0 L
Lysogeny broth (LB)	Tryptone (Oxoid LP0042)	10.0 gr
	Yeast extract (Difco 212750)	5.0 gr
	Sodium chloride (NaCl) (Baker 0278)	5.0 gr
	Thymine (Sigma T-0376)	0.02 gr
	Demi H_2O	1.0 L
Sabouraud glucose agar	Sabouraud glucose (Difco 210950	65.0 gr
	Demi H_2O	1.0 L
YMA	Yeast extract (Difco 212759)	3.0 gr
	Malt extract (Difco 218630)	3.0 gr
	Bacto peptone (Difco 211677)	5.0 gr
	D(+)-Glucose (Merck 1.08337)	10.0 gr
	Ferwo agar (Ferdiwo 10001.70.0)	20.0 gr
	Demi H_2O	1.0 L
YMA + starch	Yeast extract (Difco 212759)	3.0 gr
	Malt extract (Difco 218630)	3.0 gr
	Bacto peptone (Difco 211677)	5.0 gr
	Starch	10.0 gr
	Ferwo agar (Ferdiwo 10001.70.0)	20.0 gr
	Demi H_2O	1.0 L
YMA liquid + starch	Yeast extract (Difco 212759)	3.0 gr
	Malt extract (Difco 218630)	3.0 gr
	Bacto peptone (Difco 211677)	5.0 gr
	Starch	10.0 gr
	Demi H_2O	1.0 L
MEA (pH 7 ± 0.2)	Malt extract agar (Oxoid cm59)	50.0 gr
	Demi H_2O	1.0 L
	Malt extract (see rec28, filtrated)	400 mL

MB (pH 7 ± 0.2)	Demi H_2O	600 mL
Rec28	Malt (from brewery)	-
	Demi H_2O	Until sugar concentration is 10% (measure saccharose content according to Brix)
YPD	Yeast extract (Difco 212759)	4 gr
	Bacto peptone (Difco 211677)	8 gr
	D(+)-Glucose (Merck 1.08337)	8 gr
	Demi H_2O	400 mL
YPD + agar	Yeast extract (Difco 212759)	4 gr
	Bacto peptone (Difco 211677)	8 gr
	D(+)-Glucose (Merck 1.08337)	8 gr
	Select agar (Invitrogen 30391049)	8 gr
	Demi H_2O	400 mL
SDM JC	[RETRACTED]	[RETRACTED]
	[RETRACTED]	[RETRACTED]
	[RETRACTED]	[RETRACTED]
	[RETRACTED]	[RETRACTED]
	[RETRACTED]	[RETRACTED]

Table S3. Primers used in this study. Primers used for both amplification and sequencing of the PCR product, primers used only for amplification, and primers used only for sequencing of the PCR product in the vector after cloning are indicated in blue, green, and orange boxes, respectively. Underlined nucleotides indicate homology with the vector it is cloned into, bold nucleotides represent homology with the attL recombination sites, italic nucleotides indicate the nucleotides necessary for compatibility with a BP GateWay® Reaction, and sequences without any additional features are homologous to the targeted gene.

[TABLE RETRACTED]

Table S3. Blast results from JGI blastn against Eurotiomycetes with *AME1*.

[TABLE RETRACTED]