A bioinformatics characterization of secondary metabolism and alkyl citric acid pathway reconstruction in *Aspergillus niger* NRRL3

Sylwester Palys

A Thesis

in

The department

of

Biology

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science (Biology) at

Concordia University

Montreal, Quebec, Canada

April 2017

This is to certify that the thesis prepared

By: Sylwester Palys

Entitled: A bioinformatics characterization of secondary metabolism and alkyl citric acid pathway reconstruction in *Aspergillus niger* NRRL3.

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Chair
Dr. Patrick Gulick

_____ Examiner
Dr. David Kwan

_____ Examiner
Dr. Justin Powlowski

_____ Examiner
Dr. Isabelle Benoit Gelber

_____ Supervisor
Dr. Adrian Tsang

Approved by _____
Chair of Department or Graduate Program Director

_____ 2016 _____
Dean of Faculty

Abstract

A bioinformatics characterization of secondary metabolism and alkyl citric acid pathway reconstruction in *Aspergillus niger* NRRL3

Sylwester Palys

Secondary metabolites (SMs) from fungi have become an integral part of various scientific and medical fields as well as providing economic benefits. The genes involved in the biosynthesis of fungal and bacterial SMs are often co-localized in closely linked clusters that are generally dormant under laboratory conditions. Genomics technologies are being intensively used to identify SM gene clusters, and suggest methods to promote their expression. These efforts have already been used to discover new natural products, reveal the biosynthetic underpinnings of known products, and establish methods to eliminate toxic substances from industrially relevant organisms. An important aspect of SM gene cluster characterization is the development of methods to overproduce known compounds which are difficult to synthesize in the lab. This thesis presents the annotation of SM gene clusters in the sequenced genome of *Aspergillus niger* NRRL3. Using our cluster data as a guide to induce production of secondary metabolites through the overproduction of clustered transcription factors, we confirm the identity of the regulators that are involved in the biosynthesis of the fumonisins, TAN-1612/BMS-192548, and the azanigerones which were also used as positive controls. Additionally, we discovered three previously uncharacterized regulators of compounds whose masses correspond to the malformins, pyrophen, and the alkyl citric acids hexylitaconic acid, 2-carboxymethyl-3-hexylmaleic acid and the tensyuic acids. Selecting the alkyl citric acids for further investigation, we obtained a transcriptomic profile to define the genes of the alkyl citric acid gene cluster in *A. niger*. Using the identified compounds from mass spectrometry and NMR as well as the functional annotations of our defined cluster, I reconstruct a biosynthetic pathway for the alkyl citrates I identified. I then tested the robustness of my reconstruction by successfully predicting the accumulation of two early precursors, 2-hexylcitric acid ($\log_2$ fold change = 2.982 ± 0.15, p-value < 0.01) and its anhydride ($\log_2$ fold change = 3.792 ± 0.26, p-value < 0.01), as well as the elimination of all downstream compounds.

Acknowledgements

I would like to express my deepest thanks to my supervisor Professor Adrian Tsang who's stern and no nonsense approach to the pursuit of science has been an inspiration.

A deep appreciation is also extended to my committee members Professors Justin Powlowski, and Selvadurai Dayanandan for their willingness to act as examiners in these proceedings.

I would also like to express a heartfelt thanks to our team of postdocs who have been instrumental in my research. Thank you, Dr. My Pham for your willingness to help an old man understand the rudiments of NMR and ethyl acetate extractions. Thank you, Dr., Letian Song, for your assistance in plasmid assembly. Thank you, Dr. Lis Schwartz-Miotto, your attention to detail has served as a template for my own work; also, thanks for letting me "borrow" your solutions. Thank you, Dr. Jean-Paul Ouedraogo the Aspergillus whisperer, your knowledge and expertise in fungal growth has been indispensable. Thank you to Dr. Marcos DiFalco for helping me squirt fungal juices into the mass spectrometer. Last but not least, thank you Dr. Magdalena Kolbusz your help in getting me started into lab work and particularly cloning; it cut out several weeks of unnecessary suffering.

Additional thanks go out to the platform team, Annie Bellemare, Claudia Moen (the engine that ran the lab behind the scenes), Chy Lam (quality ingredients and love went into your preps), Sandrine Marqueteau, Yun Zheng, Tricia John (thanks for helping with PCR pipetting; I'm sure it was frustrating), Sandra Brouillard, Josée Vandal (mazel tov!), Mathieu Lavalée, Nick Geoffrion, Nadeeza Ishmael, Susan Sillaots, and Adrian Crawley-Da Costa (thanks for the MM"J" good buddy).

Big thanks are extended to Victor Levy and John Polychronoglou for providing professional references in the face of company policy. This work would not have been possible without your help.

Contribution of authors

Contributor	Statement
Sylwester Palys	Wrote the thesis; conducted the SM bioinformatics analysis. Designed the experiments relating to alkyl citric acid production and constructed alkyl citric acid over expression plasmid as well as gene deletion cassettes. Conducted experiments pertaining to alkyl citric acid production, mass spectrometry data analysis, and alkyl citric acid pathway reconstruction.
My Pham	Secondary metabolite extraction, purification, mass spectrometry data analysis and instrument operation, all NMR experiments, analysis and instrument operation. Wrote NMR methods section.
Marcos DiFalco	Mass spectrometry data analysis and instrument operation. Wrote mass spectrometry methods.
Marie-Claude Moisan	Constructed cDNA library for sequencing to obtain transcriptomic data.
Ian Duncan Reid	Transcriptomic data processing.
Erin McDonnell	Peptide localization sequence analysis.
Jean-Pierre Falgueyret	Mass spectrometry data analysis and instrument operation.
Marianne Desautels-Marissal	Plasmid assembly which included transcription factors for all non-alkyl citric acid metabolites.

Table of contents

List of abbreviations

A domain: Adenylation domain

ACP: Acyl carrier protein

antiSMASH: antibiotics & Secondary Metabolite Analysis Shell

AT: Acyl transferase

ATP: Adenosine tri-phosphate

BLAST: Basic Local Alignment Search Tool

bp: base pair

C domain: Condensation domain

CDD: Conserved domain database

CLF: Chain length factor

CoA: Coenzyme A

Cyc: Cyclization

DH: Dehydratase

DHN: 1, 8-dihydroxynaphthalene

DMAT: Tryptophan dimethylallyl transferase

E domain: Epimerization domain

ER: Enoyl reductase

FAS: Fatty acid synthase

FAs: Fatty acids

FPKM: Fragments per kilobase of transcript per million

GlaA: Glucoamylase promoter

GlaTt: Glucoamylase terminator

HPN: Hybrid PKS/NRPS

ID: Identity

KO: Gene knockout/deletion

KR: Keto reductase

KS: Keto synthase

MPT: Malonyl/palmitoyl transferase

MS: Mass spectrometry

MT: Methyl transferase

NMR: Nuclear magnetic resonance

NRPS: Non-ribosomal peptide synthetase

PCP: Peptidyl carrier protein

PKS: Polyketide synthase

PPT: Phosphopantetheine transferase transferase

PT: Product template

RNA: Ribonucleic acid

RT-PCR: Reverse transcription polymerase chain reaction.

SAT: Starter unit acyl transferase

SM: Secondary metabolite

SMURF: Secondary Metabolite Unknown Regions Finder

T domain: Thiolation domain

TE: Thioesterase

VTO: Vector transformant only

Table of figures

Table of tables

Introduction

1.1 Secondary metabolism; an overview

Secondary metabolism is the energy expenditure of a cell for metabolic processes not required for survival of the organism [1]. Bacteria, fungi and plants produce wide varieties of secondary metabolites (SMs) and any single species may possess the ability to produce many SMs from different SM classes. These classes include fatty acids, polyketides, small peptides, terpenes, sugars, and alkaloids [2-4].

With respect to the cellular functions of these molecules, a relatively small number of SM compounds have a known role. In some cases, organisms use SMs as defense mechanisms or as offensive measures against other organisms competing for space and other resources. Other cases see SMs used as signalling molecules between organisms. In addition to the advantages provided to the producing organisms, SMs have been exploited for a multitude of uses in human society [5]. This is particularly true in the pharmaceutical industry which has seen a renewed interest in biosynthesized compounds from secondary metabolism [6]

1.2 The impact of secondary metabolites on human society

The biological properties of some of the discovered SMs from fungi, plants and bacteria, sometimes referred to as natural products, have been indispensable in medicine. These include pharmaceuticals like the antibiotics penicillin and vancomycin, anti-cancer drugs like griseofluvin and cholesterol lowering agents like lovastatin [7-10]. Secondary metabolism is also used by industry for the production of food additives like caffeine from plants and citric acid and gluconic acid from fungi [11-13]. However, not all SMs are created equal; some SMs have toxic properties which can cause neural, liver and kidney tissue damage as well as death in animals [14, 15]. An overview of some of the well-known SMs including toxins is shown in Figure 1. The useful but also hazardous properties of SMs have helped prompt research into this aspect of metabolism.

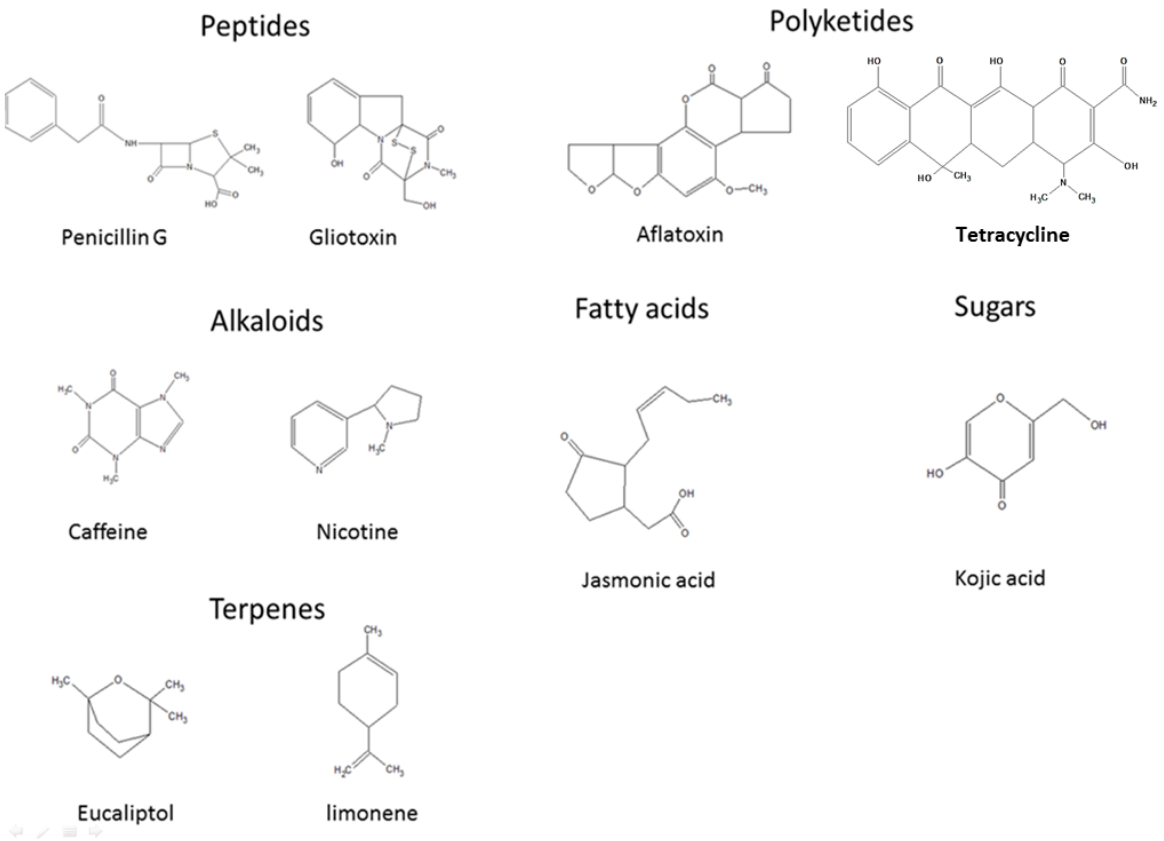


Figure 1**.** Examples of SMs from a variety of classes produced in bacteria (tetracycline), fungi (penicillin, gliotoxin, aflatoxin, kojic acid) and plants (caffeine, nicotine, limonene, eucalyptol, jasmonic acid) [16-22].

1.3 General organization of secondary metabolite genes

The genes involved in the production of SMs in bacteria and fungi are often co-localized in clusters. Each gene cluster can produce several related compounds [23]. In plants, these genes typically do not cluster; however, recent research shows that in some plant species clustering of SM genes does occur [24]. The clustered genes themselves can be highly varied but contain some important commonalities. Clusters usually encode a "backbone" enzyme that assembles the core molecular structure of a SM, "tailoring enzymes" which modify the core molecule(s) to their final product(s) and in some cases, regulatory and transporter genes facilitating cluster activation and metabolite extracellular transport respectively [25, 26] (Figure 2).

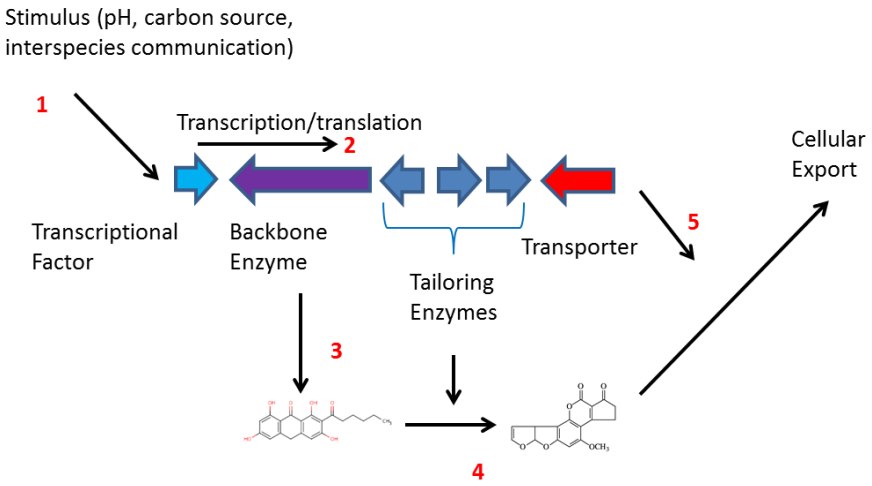


Figure 2. General structure of a SM gene cluster and the mechanism of SM production. (1) Outside stimulus induces production of a clustered transcription factor or some transacting regulator. (2) Transcription and translation of all cluster constituents. (3) Backbone enzyme begins by generating the backbone molecule. (4) Backbone molecule is modified by tailoring enzyme to generate the final products. (5) Metabolite may be exported by a clustered transporter.

Backbone enzymes in secondary metabolism can be subdivided into multiple classes defining the SM type that will be generated. These include the polyketide synthases (PKS) making polyketides, non-ribosomal peptide synthetases (NRPS) making short peptides, hybrid PKS-NRPS (HPN) combining polyketides and peptides, tryptophan dimethylallyl transferases (DMAT) making alkaloids, sesquiterpene cyclases cyclizing terpenes, and fatty acid synthases (FAS) making fatty acids [27-30].

Tailoring enzymes are separate from backbone enzymes but their encoding genes often co-localize in the same cluster. These enzymes carry out chemical reactions which add to the diversity of SMs. Examples include the linking of methyl groups to alcohol groups by o-methyltransferases and the relocation of carbon-carbon double bonds by isomerases [31, 32]. While there are many tailoring enzymes on offer, many SM clusters despite producing distinct compound products will contain common enzyme types. Enzymes like P450 monooxygenases and methyltransferases are present in numerous clusters that produce diverse classes of SMs (Table 1). Moreover, as previously mentioned, a transporter and/or transcription factor (TF) gene is present in the gene cluster which facilitates SM cellular export and cluster expression respectively [33].

Table 1. Common tailoring enzymes observed in PKS, NRPS and HPN clusters which have been experimentally characterized.

Compound	Transcription factor	Transporter	Cytochrome P450	FAD dependent protein	NAD dependent protein	Methyl/ acetyl transferase
Acetylaranotin [34]	0	1	0	0	0	1
Aflatoxin [35]	2	2	4	0	1	2
Apicidin [36]	1	1	2	0	0	1
Aspyridone [37]	1	1	2	1	0	0
Compactin [38]	1	1	1	0	0	0
Cercosporin [39]	3	1	0	3	1	1
Cyclopiazonic acid [40]	1	1	1	0	0	1
Cytochalasin E [41]	1	0	0	0	0	0
Equisetin [42]	0	1	1	0	0	1
Echinocandin B [43]	1	3	1	0	0	0
Hexadehydro-asterchome [44]	2	1	1	1	0	1
Pneumocandin B0 [45]	1	1	2	0	0	1
Pseurotin A [46]	0	0	1	0	0	1
Pyranonigrin E [47]	1	0	0	1	1	1

[1]FAS alpha subunits not shown but are present in aflatoxin and apicidin clusters. *Production of Gluconic Acid by Some Local Fungi.*

1.4 The backbone enzymes

1.4.1 PKS structural organization

Most PKS enzymes are multi-domain proteins or protein complexes. To be considered a canonical PKS, the enzyme must contain a minimal set of domains. The minimal set includes the keto-synthase domain (KS), the acyl-transferase domain (AT) and the acyl-carrier protein domain (ACP) [48]. In addition to this minimal set, reducing or optional domains may be present. These help generate diversity in polyketide core molecules through different combinations of ketone, alcohol and alkene bond reduction but are not required to produce a polyketide [49]. The reducing domains include the dehydratase domain (DH), the keto-reductase domain (KR), and the enoyl reductase domain (ER). In some cases, methyltransferase (MT)

domains (adding methyl groups) and thioesterase domains (performing terminal cyclization and polyketide-enzyme release reactions) may also be present [50] (Figure 3).

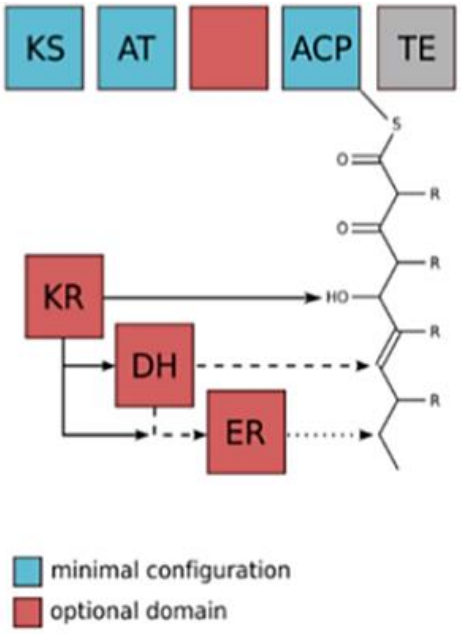


Figure 3. General structure of a PKS backbone enzyme. The ketone, alcohol and alkene bond reduction reactions are depicted. Blue and red squares represent minimal and reducing/optional domains respectively. Modified from Kehr, J.C., et al [51].

There are three generally recognized types of PKS enzymes. These are designated type I (further broken down into iterative and modular), type II and type III [52, 53]. Modular type I PKSs are multi-protein production lines with each protein or module containing various combinations of minimal and reducing PKS domains. Each module is typically used only once during the elongation process to form the complete backbone molecule. Type I modular PKSs are the common form found in bacterial species.

Type I iterative PKS enzymes are single unit large proteins containing all minimal domains and sometimes reducing domains. Each domain can be used multiple times in an iterative elongation process creating a growing polyketide chain [53]. Following each elongation iteration adding two carbons from a malonyl-CoA extender unit, the reducing domains reduce ketone groups in the polyketide to alcohols, to double bonds or finally to single bonds [54]. The iterative type I PKSs can be further subdivided into subclasses based on which reducing domains

are present. The PKSs containing just the minimal set of domains (KS, AT and ACP) and sometimes a TE or MT domain are considered non-reducing. The non-reducing type I iterative PKSs can also contain starter unit acyl transferase (SAT) and product template (PT) domains. The SAT domain allows a non-reducing PKS to use different starting compounds in the first iteration of polyketide production while the PT domain facilitates polyketide cyclization [55]. Structures resulting from the cyclization reactions of PT domains may to some extent predicted by the size of the catalytic pocket within the domain [55]. Type I iterative PKSs that have reducing domains make up the other three subclasses. A PKS is highly reducing when all reducing domains (DH, KR and ER) are present, partially reducing when some reducing domains are present (KR and/or DH) [50] or a 6-methylsalicylic acid synthase PKS when a DH and KR domain is present but not MT or TE domains [56, 57].

These different classes of type I iterative PKS enzymes will produce polyketides with differing structures and functional groups. The highly reducing PKS enzymes generally produce backbone compounds where most or all keto groups are fully reduced to alkane single bonds [58]. For partially reducing PKSs, the backbone compounds are partially reduced polyketides containing alcohol groups and alkene carbon-carbon bonds [59]. Non-reducing PKSs produces backbone compounds that are "true" polyketides containing only keto groups on the hydrocarbon chain. Additional downstream modifications also tend to produce aromatic/cyclic configurations of these polyketides. Lastly, the 6-methylsalicylic acid synthase PKS produces only 6-methylsalicylic acid, a small tetraketide used in multiple SM pathways [56, 59].

The type II and type III PKSs are all single domain proteins and like their multi-domain counterparts can in some cases reuse those domains iteratively to produce polyketides [52, 60]. In type II PKSs all minimal and reducing domains may be present, but these exist as protein complexes where each domain is contained in a separate protein. With the type III PKSs, sometimes referred to as chalcone synthases, only a single protein is produced which contains a KS domain. These KS domain proteins exist as a homodimer of two identical KS domain containing proteins. Overall, in filamentous fungi PKS enzymes are generally of the type I iterative form. In some species of fungi the type III PKS may also be present but these enzymes are more commonly found in plants. In the case of the type II PKSs, these can be found in bacterial species [53, 61].

1.4.1.1 General PKS assembly of polyketides (PK) in fungi

The process of assembly of the polyketide begins by activation of the ACP domain in a PKS by a phosphopantetheinyl transferase (PPTase) enzyme transferring a phosphopantetheinyl arm to a conserved serine residue. An acyl starter unit, usually an acetyl-CoA, is loaded onto a conserved cysteine residue of the AT domain which then transfers the unit to the KS domain aided by the ACP. Next, the AT domain catalyzes the loading and subsequent transfer of a malonyl-CoA extender unit to the phosphopantetheinyl arm of the ACP domain. Other extender units may also be used like the methylmalonyl-CoA [62]. The KS domain then carries out a Claisen condensation reaction losing CO_2 to form a diketide with its loaded starter and the extender unit on the ACP domain. In highly reducing PKSs and partially PKSs the polyketide is reduced by the actions of DH, KR and/or ER domains (Figure 3). In some PKSs, MT domains can also add methyl groups to the growing polyketide. The polyketide is then transferred back to the KS domain and the process continues with another extender unit until a specific length is reached. In some cases, chain length is determined by a chain length factor (CLF) subunit of the KS domain. At the end of the last iteration, the polyketide releases spontaneously or via a TE domain if present. In many non-reducing PKSs the SAT domain selects a starter unit other than acetyl-CoA. For example, a 6-carbon (C6) fatty acid is used to initiate aflatoxin biosynthesis. Unlike most bacterial PKSs, in fungi the iterative PKS can use both the minimal and reducing domains multiple times. The sequence of reducing domains can also be varied. For instance, in the first iteration the keto group may be fully reduced to an alkane but in the next iteration the keto group may only be reduced to an alcohol or not reduced at all. While much is known about the iterative process of polyketide assembly in fungi, prediction of the number of iterations to deduce chain length remains elusive. Only general statements can be made concerning the reduction state of the backbone molecule [17, 63-65]. A general mechanism of backbone compound biosynthesis is presented in Figure 4.

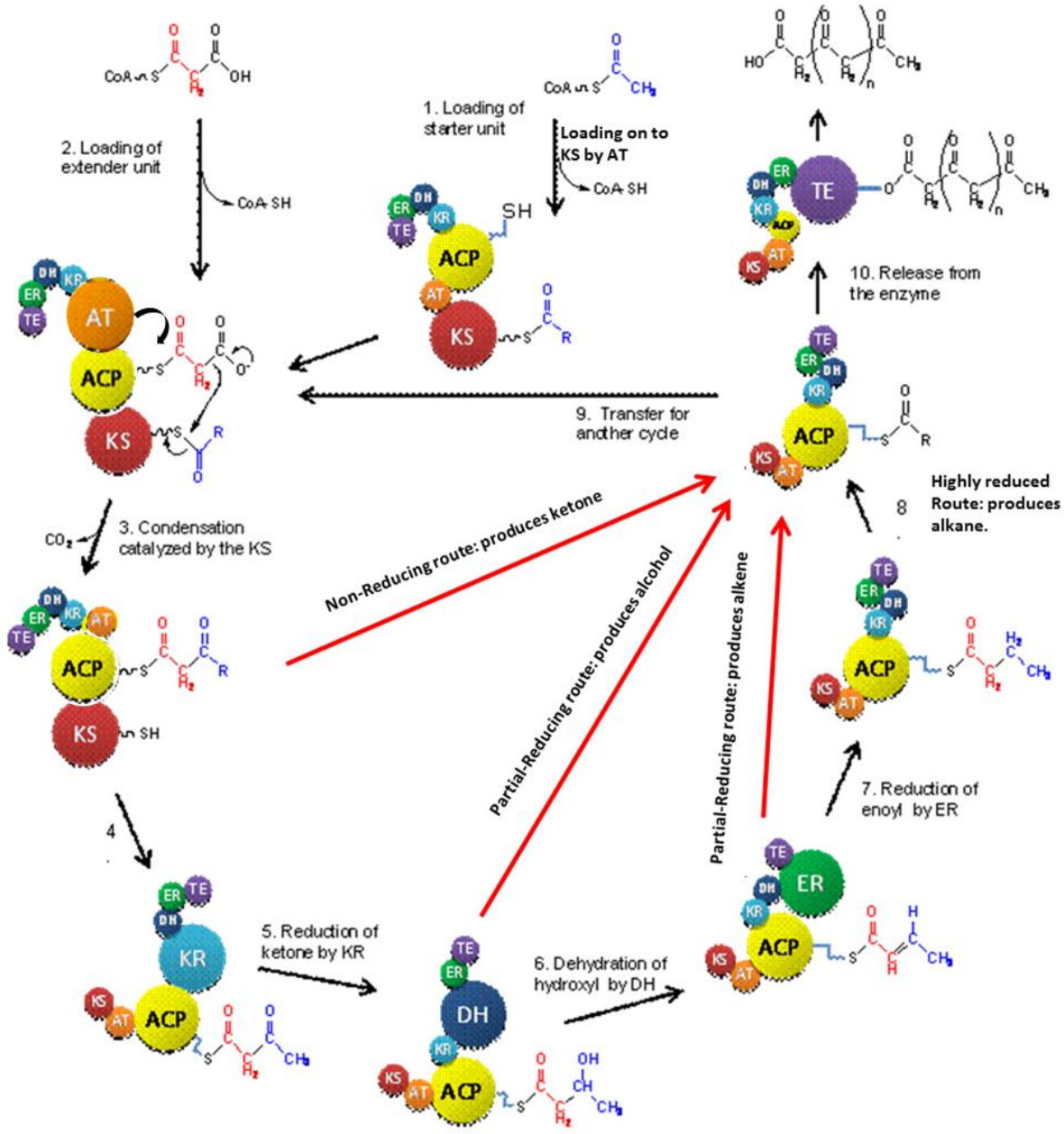


Figure 4. Schematic of polyketide assembly. The SAT and PT domains are omitted but the action of the SAT domain in non-reducing PKSs occurs in step 1 loading a non-acetyl-CoA starter unit. The point at which the PT domain acts is currently unclear. Modified from Rasmus J. N. F. [65].

1.4.2 NRPS structural organization

Non-ribosomal peptide synthetases in fungi produce amino acid based SMs in an assembly line fashion without the use of a ribosome. The NRPS selects amino acids and incorporates them into a growing peptide chain. These enzymes comprise multiple fused modules, each composed of several domains (Figure 5). Unlike ribosomal peptide synthesis, which can also produce SM peptides [66], NRPSs do not require RNA template for biosynthesis and can also incorporate non-proteinogenic amino acids into the peptide chain such as D-amino acids. Three types of NRPS systems have been defined: The linear form where all domains are used once during assembly, the iterative form where sites can be reused cyclically during assembly, and the non-linear form where some modules and/or domains in the same NPRS can be reused [50].

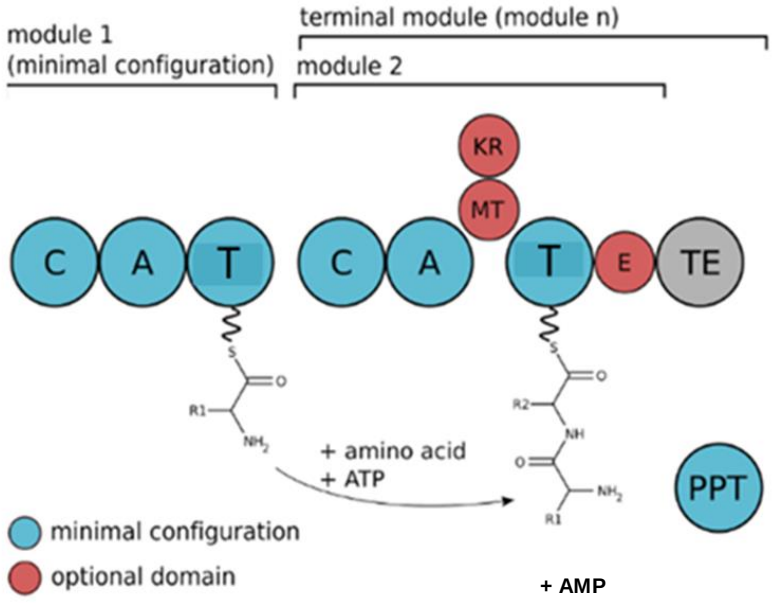


Figure 5. General structure of an NRPS backbone enzyme. Amino acid adenylation is depicted. Blue and red circles represent minimal and reducing domains respectively. Modified from Kehr, J.C., et al [51].

Like in the PKS backbone enzymes, NRPSs have minimal domains. There are three minimal domains in an NRPS: the adenylation domain (A domain), a peptidyl carrier protein or

thiolation domain (PCP or T domain), and a condensation domain (C domain). Each A domain selects a single amino acid and then activates it by adenylation from ATP hydrolysis. The T domain acts as a placeholder for the peptide chain and is similar in function to the ACP domain in PKSs. Lastly, the C domain carries out the peptide bond reaction joining amino acids to the growing peptide [67].

The NRPS enzymes like PKS enzymes can be further broken down into linear, iterative and non-linear NRPS forms. Unlike the PKS enzymes the subclass distinctions of NPRS enzymes is not so clear but some general inferences can be drawn. The general organization of domains in linear NRPSs is made up of sequential A-T-C modules: A-T-C--A-T-C--A-T-C… etc. In iterative NRPSs, each domain is reused and the generated SM peptide is composed of repeating amino acid subunits; these systems are generally considered linear with iterative function near the C-terminus. The C-terminus of iterative NRPSs may contain multiple T and/or C domains presenting as C-A-T--C-A…T-T-C or A-T-C--A-T…C-A-T-C-T-C-T-C. With non-linear systems, domains can be used once or multiple times which makes prediction of the SM difficult. The domain organization of these systems is highly varied for example: A-T-E-C--A-C-A-C-A-C--A-T-E-C-T-C-T and C-A--T-A-T-C--A-T-E-C--A-T-E [2, 68].

Minimal domains tend to appear together as a module, fused together in fungi while in bacteria they can be separate entities [69]. Optional domains may also be present within NRPS modules generating further product diversity. These domains can include some of the domains present in PKSs (MT, Te, and reductases) but also domains unique to NRPSs like epimerization (E) domains converting amino acids from the L to D configuration and cyclization (Cyc) domains catalyzing cyclization reactions in the nascent peptide chain [50, 70] (Figure 5).

Some NRPS enzymes hold to the rule of collinearity. This rule states that the amino acid sequence of a non-ribosomal peptide product is determined by the structure and sequence of the modules and their domains from its parent NRPS. The collinearity rule generally applies to the linear NRPSs class only and can allow prediction of peptide length and composition from the order of the domains, the amino acid substrate specificity of the individual A domains in each module and the order of the modules themselves [50, 70]. This rule has been used to predict the structure of novel compounds from NRPS A domain sequences. For example, orphamide A was predicted from the protein sequence of A domains, and subsequently isolated and verified [71].

1.4.2.1 General NRPS assembly of non-ribosomal peptides (NRP) in fungi

In linear models, the initial step in NRP synthesis is the selection of an amino acid by the A domain of the first NRPS module. The selected amino acid is converted into an aminoacyl-AMP derivative by the action of the A domain and ATP hydrolysis. The adenylated amino acid is transferred to the T domain, which is activated by the addition of a phosphopantetheinyl (PPT) arm by a PPTase; much like the ACP domain in PKS enzymes. Next the selection of another amino acid by the A domain of the second module is carried out and subsequently transferred to the T domain of that module. The C domain of the second module carries out the amide bond formation between the first T domain and the second. The reactions proceed in the same manner across subsequent modules until peptide release sometimes aided by a TE domain. Optional domains can also be present. Epimerization domains can carry out reconfigurations of stereo centers while Cyc domains facilitate cyclization reactions in the nascent peptide (Figure 6). This sequence of events accounts for why most linear NRPS initiation modules lack a C domain since the condensation reaction between the first and second modules is catalyzed by the C domain in the second module [72].

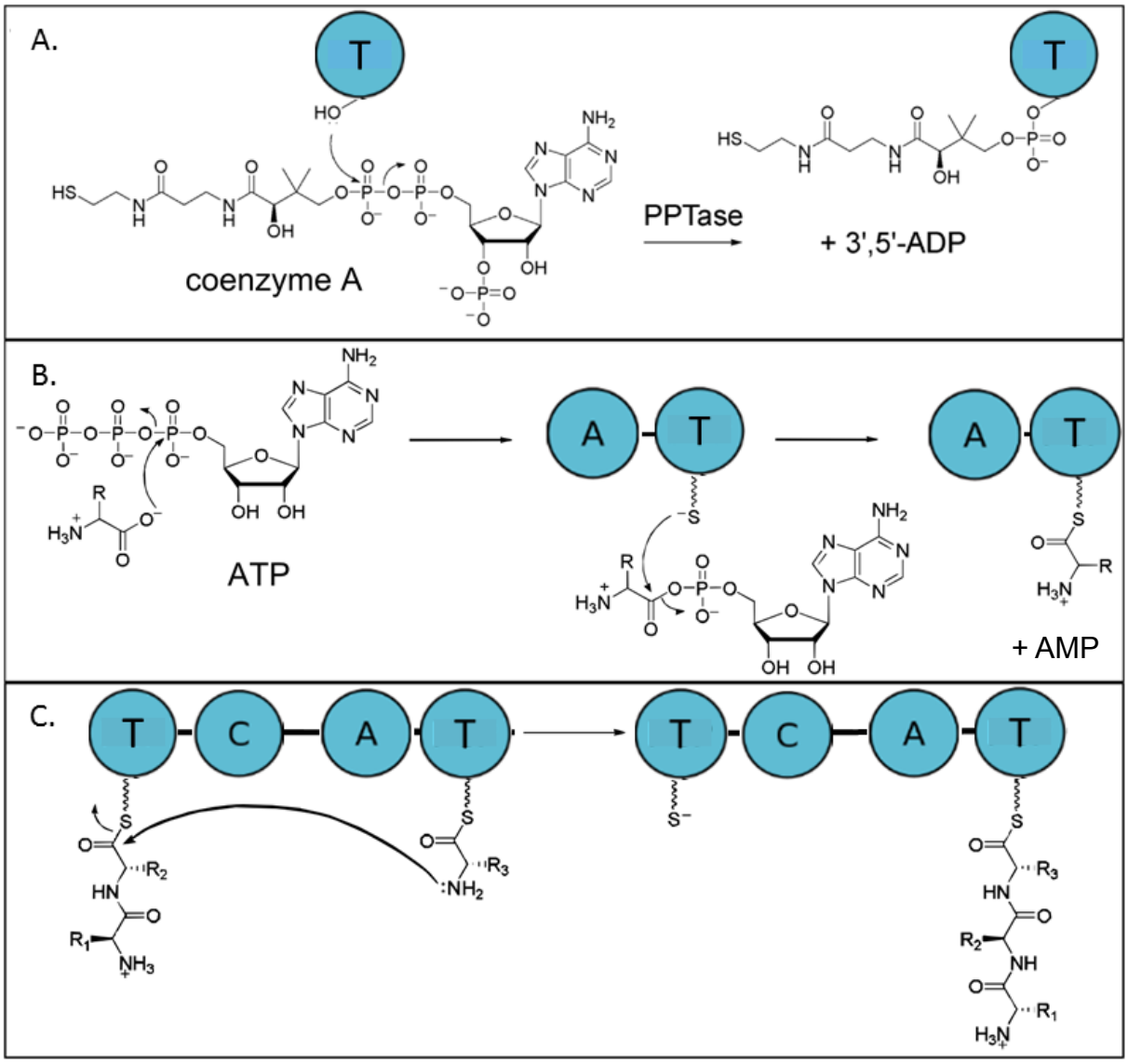


Figure 6. NRPS peptide assembly overview. Activation of the T domain by a PPTase through phosphopantetheinyl transfer (A). The A domain selects an adenylated amino acid and transfers it to the T domain (B). The C domain forms a peptide bond between one amino acid and another amino acid selected by a downstream module in the same way as the initial module (C). Modified from 2013.igem.org/wiki/index.php?title=Team:Heidelberg/NRPS&oldid=353123 [73]

1.4.3 HPN, NRPS-like and PKS-like structural organization

Hybrid PKS-NRPS enzymes in filamentous fungi can take two forms. One form has both a PKS and NRPS enzyme produced by separate genes in the same cluster. These backbone enzymes ultimately fuse their products into a single hybrid polyketide-non ribosomal peptide. The HPN generating pneumocandin is an example of this form, fusing a polyketide with amino acid residues [45, 50]. The other HPN form derives from a single gene that encodes both a PKS and an NRPS module. The HPN enzyme that produces pseurotin and equisetin are two examples of this form [42, 46, 50] (Figure 7). Fused HPNs typically start with a PKS module followed by a single NRPS module with the domains C-A-T. The PKS portion of the HPN also tends to be a highly reducing class, but partially reducing forms are also known. Production of hybrid molecules proceeds as previously described for the polyketide and peptide components. The attachment of the polyketide chain to the selected amino acid is accomplished by a C domain of an NPRS module [74].

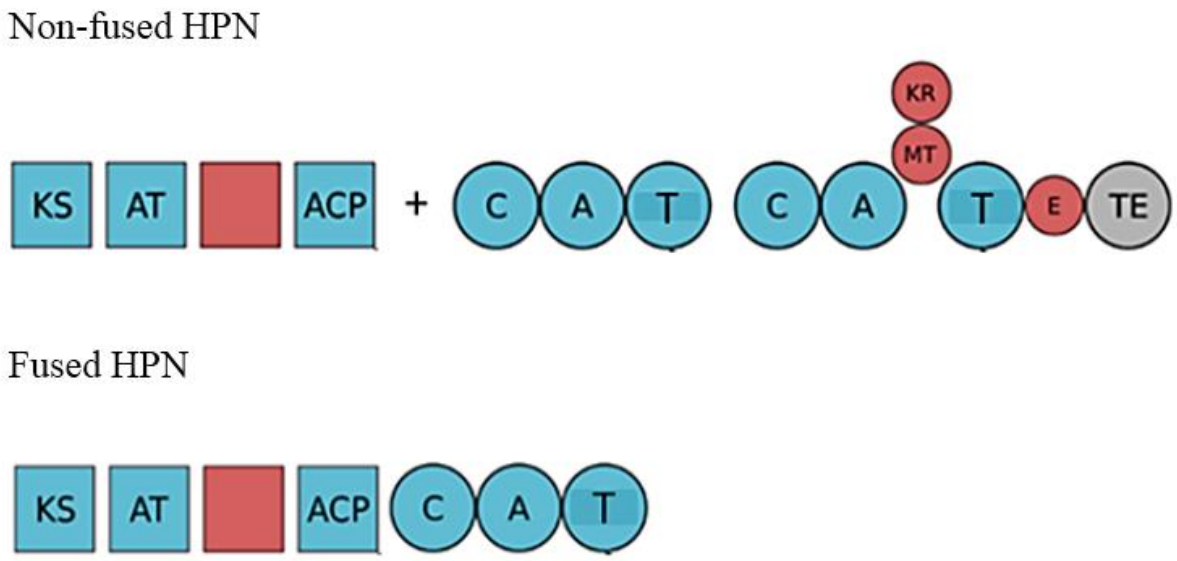


Figure 7. General structure of both HPN backbone enzyme forms. Blue and red represent minimal and reducing domains respectively. Squares and circles respectively represent the PKS and NRPS domains. Modified from Kehr, J.C., et al [51].

Genes encoding type 1 iterative PKS and all NRPS class enzymes that are missing one or more minimal domains are found in fungi [75, 76]. The NRPS enzymes missing minimal

domains are referred to as NRPS-like while PKS enzymes missing at least one minimal domain are referred to as PKS-like. The NPRS-like enzymes can still participate as backbone enzymes in SM biosynthesis. This has been recently demonstrated in the production of the dipeptide atromentin fusing two modified tyrosine residues [77]. Unlike the type III PKS and NRPS-like backbone enzymes, PKS-like enzymes have no known function to our knowledge but may be active from trans-acting single domain surrogates like in the case of trans AT domains in bacteria [78]. Their genes may in fact be remnants of full-length PKS or HPN genes whose products contain at least the minimal domains. For example, a PKS-like gene found in the genome of *A. niger* ATCC 1015 appears to be remnant of the PKS gene known to be involved in ochratoxin production in *A. niger* CBS 513.88 [75].

1.4.4 Fatty acid synthase structural organization and assembly

The FAS enzymes of filamentous fungi can be involved in both primary and secondary metabolism. However, distinct forms of FASs are used for each type of metabolism [28]. The domain complement of FAS is very similar to PKS enzymes with an additional PPT domain that for PKS enzymes is usually a separate enzyme encoded by a separate gene [79]. Only one known iterative PKS from *Saccharopolyspora erythraea* appears to violate this rule [80]. The malonyl/palmitoyl transferase (MPT) domain in fungal FAS enzymes is another domain not present in PKS enzymes. Its function is to transfer a hydrocarbon chain back to a coenzyme A (CoA) following the completion of chain elongation [80, 81]. Lastly, unlike the PKS, the fungal FAS exists as a heterodimer consisting of an alpha subunit with domains MPT-ACP-KR-KS-(PPT) and a beta subunit with domains AT-ER-DH-MPT [80].

Fatty acid assembly is analogous to polyketide assembly using acetyl-CoA starter and malonyl-CoA extender units to build hydrocarbon chains [82]. The FAS can also incorporate a variety of starter units like the non-reducing PKS [83]. A notable difference between these two classes of backbones is that chain elongation in FASs typically occurs after full reduction of the keto group from the malonyl unit to a single carbon-carbon bond while in PKSs extension can occur with or without keto reduction [84].

1.4.5 Structure and function of tryptophan dimethylallyl transferases and sesquiterpene cyclases.

The DMAT and sesquiterpene cyclase enzymes in contrast to the preceding backbone enzymes contain only one catalytic/biosynthetic domain. Despite being single domain enzymes, the genes coding for these two backbone classes are also found in SM gene clusters like PKS and NRPS genes (68, 69). In terms of core molecule production, the DMAT enzymes produce alkaloids while sesquiterpene cyclases produce cyclized terpenoid compounds. Biosynthesis by DMATs and sesquiterpene cyclases differ in the backbone metabolites they generate. The DMAT generates a modified tryptophan residue (attached dimethylallyl group), while the sesquiterpene cyclase cyclizes an assembled sesquiterpenes such as farnesyl in tricothecene biosynthesis [85, 86].

1.5 Bioinformatics approaches to identifying and annotation SM clusters

With the advent of genomics, many aspects of annotation have become automated. Today, automated annotation includes a host of software applications to identify SM clusters. These applications use common features from experimentally characterized SM clusters to define new clusters in sequenced genomes. For instance, software such as SMURF and antiSMASH locate backbone enzyme genes in genome sequences and work outward adding neighbouring tailoring genes found in experimentally verified clusters to the cluster being defined if they appear nearby [87, 88]. Agreement between various SM cluster defining software can be problematic and ultimately requires manual input to refine the cluster edges (Figure 8) [27]. As a first approach, however, software such as SMURF and antiSMASH can reduce the time needed to locate and define clusters compared to solely manual methods [87, 88]. Additionally, applications such as Sybil use orthology of SM gene clusters between sequenced genomes of different species to help guide manual definition of their boundaries. Using Jaccard clustering methods (BLASTP queries of all proteins from all the selected genomes against each other) Sybil identifies orthologues between species. This is based on the premise that all the genes contained in a specific SM cluster shared between different species (%ID's between species > 80%) will produce the same SM and be fully retained from species to species. For instance, two clusters from two species matched up in Sybil show that seven of the genes matchup between them (%ID for each gene >80%) genes outside these seven do not match.

Based on this observation the cluster in both species has only seven genes and the boundaries can now be drawn [89]. Sybil provides graphical results to help reveal possible clusters edges defined at junctions where the orthology ends in all matched genomes (Figure 8).

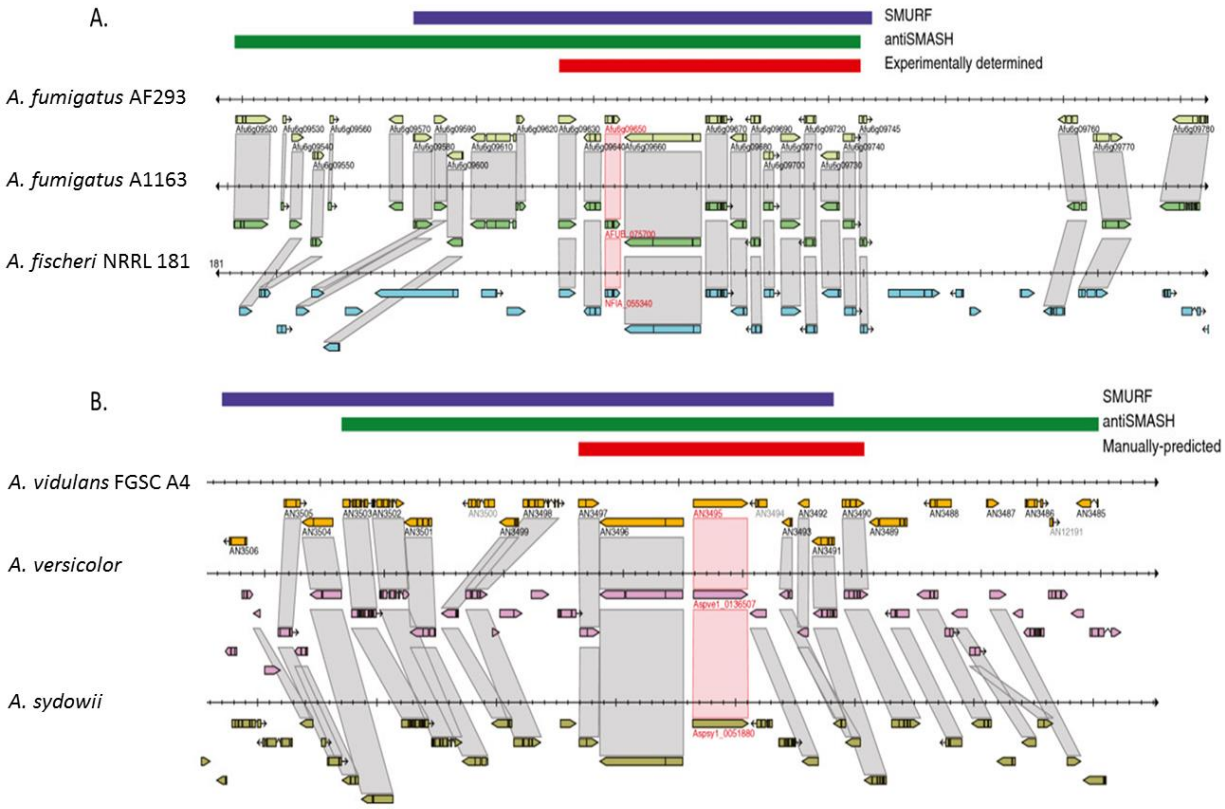


Figure 8. Sybil display of the experimentally determined cluster producing gliotoxin in *Neosartorya fischeri* compared to other genomes (A). The complete cluster appears in two *A. fumigatus* strains (top and middle) with the junctions of non-orthologous and orthologous genes (either ends of the red bar) defining the experimentally determined edges of the gliotoxin cluster. The Sybil result in B shows an unknown cluster predicted using the same method that showed proof of concept in A; the edges are again defined at orthologous/non-orthologous junctions ends of the red bar) across all matched genomes. Taken from Inglis, D.O., et al [27]

1.5 Filamentous fungi produce thousands of "orphan compounds" and contain a wealth of "cryptic" SM clusters

Filamentous fungi are prolific producers of natural products with identified compounds numbering in the thousands [90]. The genes and/or gene clusters responsible for the production of many of these compounds are unknown. Compounds for which the biosynthetic processes are unknown are referred to as "orphan compounds" [91]. Many of these orphan compounds have useful properties and determining their genetic underpinnings could allow their overproduction. For example, the tensyuic acids and hexylitaconic acid are orphan compounds discovered in *A. niger* that have a number of useful biological properties. Hexylitaconic acid, known since the 1980's has been shown promote plant growth [92] while tensyuic acid C has shown potential as an anti-parasitic and antibiotic compound [93, 94]. Determining the biosynthetic pathway for these compounds could help devise strategies for their overproduction. This has been shown recently in *Cercospora nicotianae* where knowledge of the biosynthetic pathway for cercosporin, a polyketide, was used to produce mutants which accumulate various precursor compounds [95]. The ability to overproduce these useful SMs can reduce the cost of their production by reducing the need for purification and improve their adoption in health, environmental and industrial applications.

A key step in pathway reconstruction is to locate the parent gene cluster of a compound(s) to determine the enzymes involved. Many of the clusters annotated in the sequenced genomes of fungi have unknown functions. These clusters are referred to as "cryptic" [96]. By annotating the set of SM clusters in a particular species, the search for the genetic underpinnings of particular orphan compounds can be narrowed down. Then, through analysing the tailoring and backbone genes within clusters, their functions can be used to match a particular orphan compound with its cryptic cluster. In a recent work, a P450 monooxygenase and an o-methyltransferase enzyme was used as a guide to a specific non-reducing PKS containing cluster responsible for the production of a set of orphan polyketides referred to as the kotanins in *A. niger* [97].

1.6 *Aspergillus niger* NRRL3 as a model fungus for SM research

With 228 of its 13870 genes having gene onotology references to secondary metabolism [27], *A. niger* has good potential for identifying gene clusters involved in the production of orphan compounds as well as for novel compound discovery. Compared to other well studied Aspergilli, the number of SM clusters in *A. niger* (84) exceeds that annotated in the genomes of *A. nidulans* (66), *A. fumigatus* (34), and *A. oryzae* (73) making it a good candidate for the study of SM production. *Aspergillus niger* and related species are also known to produce over 140 SMs 68 of which have not been tied to their gene clusters (Table 2.) [98]. This leaves the biosynthetic pathways for many of these orphan compounds available for characterization.

Table 2. Secondary metabolite compounds isolated from *Aspergillus niger*.

Polyketides	Orphan	Non ribosomal peptides	Orphan
* 8′-O-Demethylisonigerone		Malformin A1	X
* 8′-O-Demethylnigerone		Malformin A2	X
* 10,10′-Bifonsecin B		Malformin B1a	X
* Aurasperone A		Malformin B1b	X
* Aurasperone B		Malformin B2	X
* Aurasperone C		Malformin B3	X
* Aurasperone D		Malformin B4	X
* Aurasperone E		Malformin B5	X
* Aurasperone F		Malformin C	X
* Aurasperone G		Cyclo-l-Ala-l-Leu	X
* Asperpyrone A		Cycloleucomelon	X
* Asperpyrone B		Atromentin	X
* Asperpyrone C			
* Dianhydroaurasperone C		Terpenes	Orphan
* Flavasperone		Asperrubrol	X
* Flaviolin		Yanuthone A	
* Fonsecinone A		Yanuthone B	
* Fonsecinone B		Yanuthone C	
* Fonsecinone C		Yanuthone D	
* Fonsecinone D		Yanuthone E	
* Fonsecin (TMC-256B1)		Nafuredin	X
* Fonsecin B (TMC-256B2)		Asperenone	X

* Fonsecin monomethyl ether		22-Deacetylyanuthone A	
* Isoaurasperone A		1-Hydroxyyanuthone A	
* Isonigerone		1-Hydroxyyanuthone C	
* Nigerone			
* Rubrofusarin		Alkaloids	Orphan
* TMC-256A1		Aspernigrin A	X
* TMC-256C1		Aspernigrin B	X
** Bicoumanigrin		Pyrophen	X
** Demethylkotanin		Tensidol A	X
** Orlandin		Tensidol B	X
** Kotanin		Nygerone A	X
** 7-Hydroxy-4-methoxy-5-methylcoumarin		Nigragillin	X
Hexahydroxy-5H,6H,7H-benzopyrene-1,11-dione	X	Nigerazine A	X
3',4',5,7-Tetrahydroxy-8-methoxy isoflavone	X	Nigerazine B	X
4,9-Dihydroxyperylene-3,10-quinone	X	Diketopiperazine dimer	X
8-Hydroxygenistein	X	Asperic acid	X
Agglomerin [99]		Aspernigerin	X
Asnipyrone A	X	Asperazine	X
Asnipyrone B	X		
Asperyellone	X	PK/NRP hybrids	Orphan
Azanigerone A [100]		Pyranonigrin A	
Azanigerone B [100]		Pyranonigrin B	
Azanigerone C [100]		Pyranonigrin C	
Azanigerone D [100]		Pyranonigrin D	
Azanigerone E [100]		Pyranonigrin E [47]	
Azanigerone F [100]		Ochratoxin A	
BMS-192548		Ochratoxin α	
Carlosic acid [99]		Ochratoxin β	
Differenol A	X	Nigerasperone A	X
Fumonisin B2		Nigerasperone B	X
Fumonisin B4		Nigerasperone C	X
Funalenone	X	Aspergillin	X
Iso-T-2 toxin	X		
Orobole	X	Fatty acids	Orphan
Nigerapyrone A [101]	X	2-Carboxymethyl-3-hexylmaleic anhydride	X
Nigerapyrone B [61, 101]	X	2-hexylcitric acid	X
Nigerapyrone C [101]	X	Hexylitaconic acid	X

Nigerapyrone D [101]	X
Nigerapyrone E [101]	X
Nigerapyrone F [101]	X
Nigerapyrone G [101]	X
Nigerapyrone H [101]	X
Nigerloxin	X
Sorbic acid	X
TAN-1612 [102]	

Tensyuic acid A [93]	X
Tensyuic acid B	X
Tensyuic acid C	X
Tensyuic acid D	X
Tensyuic acid E	X
Tensyuic acid F	X

Other	Orphan
(+)-Parasorbic acid	
2-Phenylethanol	
2-Methylene-3-hexyl-butanedioic acid	X
3-Methyl-8-hydroxy-4-decanoate	X
4-Hydroxymandelic acid	
d-Galactonic acid	
Glyoxylic acid	
Glutaric acid	
Glycolic acid	
Hydroxypyruvic acid	
Fumaric acid	
p-Methoxyphenylacetic acid	X
Phenoxyacetic acid	X
Phenylacetic acid	X
Pisolithin B	X

[1]Table modified from Nielsen, K.F., et al [98].
[2]Orphan compounds are marked with an X.
[3]Compounds without associated reference were sourced from Nielsen, K.F., et al [98].
[4]Where available, compounds have been categorized according to Nielsen, K.F. et al [98] or from literature searches. All other compounds categorized based on typical backbone molecular structure and composition for each class.
* Compound is produced by *albA*/*fwnA* PKS and its associated tailoring enzymes based on literature and core structure similarity [103-105].
** Compound is produced by the *ktnA* PKS cluster based on literature and core structure similarity [97].

1.7 Expression, analysis and biosynthetic pathway reconstruction of SM clusters

While in silico cluster definition can serve as an initial approximation of the genes within a given cluster, ultimately experimental verification is required to test the robustness of predictions. Expression data in the form of RT-PCR or RNA-seq can be used to support cluster predictions [100, 106]. Transcriptomic data from RNA-seq may also be useful in identifying non-clustered genes involved in secondary metabolite production since it takes into account gene expression data of the entire genome.

The expression of SM clusters in nature can occur from a wide variety of internal and external stimuli [107]. Under laboratory conditions however, many SM clusters remain silent and activating their expression can prove challenging [108]. Attempts to induce cluster expression have met with some success through methods such as overexpressing a clustered transcription factor or a global transcription factor [100, 109], modification of the epigenome [110], or heterologous expression in other hosts including *Saccharomyces cerevisiae, A. nidulans and A. oryzae* [102, 111]. Developing strategies to solve the expression problem is an active area of research that may prove beneficial in the discovery of novel compounds and serve as a means to overproduce valuable known compounds. Alternately, the determination of the clusters responsible for producing known toxins may provide a means for their removal from industrially relevant organisms like *A. niger* used in the production of food additives [12]. Malformin for instance, which causes cell-cycle arrest in mammalian cells, has been known to be produced by *A. niger* since at least the 1950's [112] but the determination of the underlying genes responsible for its production remain elusive.

1.8 Rationale for pursuing research in fungal secondary metabolism

Research into fungal secondary metabolism is of vital importance not only to human health but also to the various chemical products used in human society. Understanding the mechanisms of SM production has the potential to reduce health risks in microbial food production, develop new lifesaving drugs (particularly as antibiotic resistance is becoming a health issue), and increasing production of those drugs and other commodities. These are just some of the aspects that highlight the need for additional research into this field of biology.

1.8.1 Toxic substances can be produced from organisms used in industry for food production

Toxins produced by fungi can pose public health risks, particularly when fungi are used in industry [15, 113]. The elimination of these toxins may mitigate some of the risks. For example, the filamentous fungus *Aspergillus niger* is grown in large fermentation vats and its metabolic processes are exploited for the production of a variety of food additives like citric acid [12]. Some strains however also produce a variety of unwanted toxins such as fumonisin and ochratoxin. Given the consumption of food additives produced in *A. niger*, compounds like fumonisin and ochratoxin pose potential health risks to people and livestock [113]. With the prior annotation of the fumonisin gene cluster in *A. niger* and other species of filamentous fungi, the production of fumonisins may potentially be eliminated by targeted gene deletion [114]. However, while true for some toxins, many toxins produced in *A. niger* as well as other filamentous fungi used in industry have no known genetic underpinnings and still pose potential risks to public health.

1.8.2 Secondary metabolism and antibiotic resistance

The discovery of novel compounds is becoming an increasing priority in university research particularly with respect to antibiotics since the pharmaceutical industry has begun diverting funds to more profitable areas [115, 116]. Over the last few decades two main issues have been occurring in the public health field that are prompting investigations into the current state of available antibiotics. First, the number of new antibiotics being discovered has been falling [116] and second the rates of bacterial resistance to current antibiotics is on the rise [117]. In fact, bacterial species like methicillin resistant *Staphylococcus aureus* are now resistant to many antibiotics resulting in several difficult to treat diseases [118]. These two aspects of an unfolding health crisis are generating a pressing need to discover new compounds. Current research in this area is not keeping pace with the build-up of resistance by microorganisms to existing antibiotics [119].

One potential avenue to finding new antibiotics is to examine the secondary metabolism of microorganisms. Many of the antibiotics discovered in the past, like penicillin, are produced

from the secondary metabolism of fungi [120]. With the advent of genomics, recently sequenced genomes of various fungi reveal that they possess genes capable of generating numerous new secondary metabolites, some of which could be potential antibiotics [121]. This fact alone illustrates that research into secondary metabolism is a potential "gold mine" for not only generating new antibiotics, but also other novel compounds.

The threat of not just antibiotic resistance but other medical issues as well, is bringing pressure to bear on researchers to discover novel compounds. One advantage of using filamentous fungi for novel compound discovery is the number of orphan clusters within their genomes. To put this into perspective, the genome of *A. niger* contains at approximately 80 SM clusters (strain dependent) but only about 10%, from diverse SM classes, have had their metabolic products experimentally determined [27, 47, 97, 99, 100, 102, 122-124]. Figure 9 shows the variation in chemical structure of some of the known products generated by *A. niger* including orphan compounds.

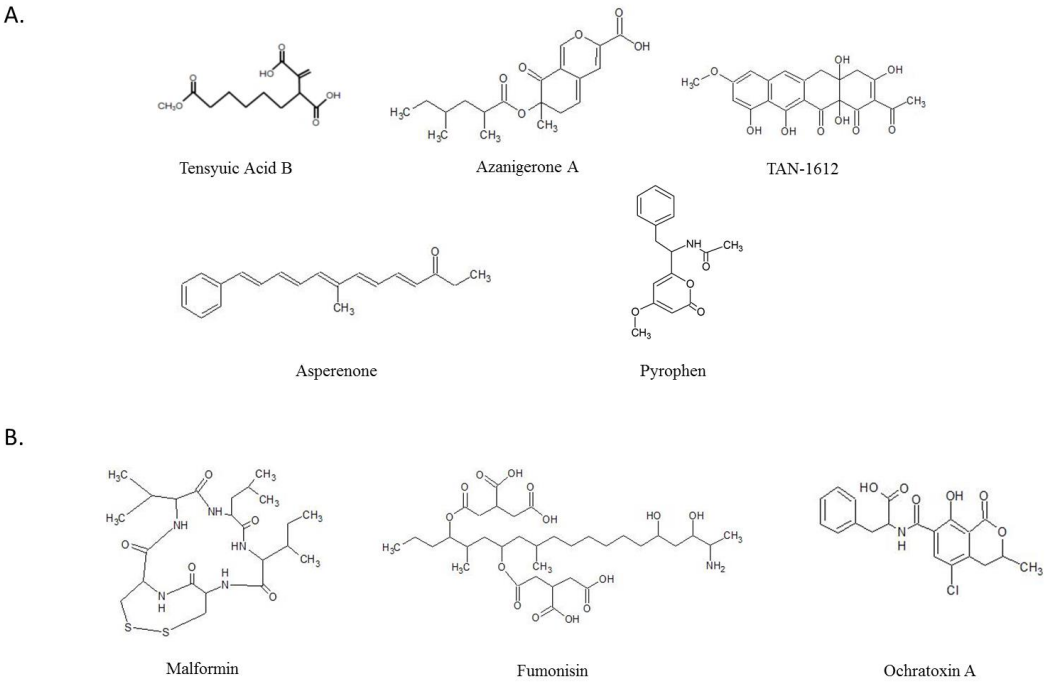


Figure 9. Secondary metabolites (A) and toxins (B) produced by *Aspergillus niger*. Orphan compounds include malformin, tensyuic acid B, asperenone and pyrophen.

1.8.3 Manipulation of SM pathways can overproduce desired products

Many secondary metabolites are complex molecules and consequently their chemical synthesis can be very expensive. For this reason, having organisms like *A. niger* produce desired SMs can substantially cut costs of production [125]. Enter metabolic engineering; using the knowledge from experimentally verified pathways, researchers are now able to manipulate these pathways on a genetic level towards accumulating desired compounds. Varieties of methods and techniques have been employed to accomplish this goal. These include pathway gene deletions resulting in terminations at particular biosynthetic steps, elimination of competing pathways by gene deletion to increase the resources available for the desired pathway and *de novo* pathway design by combining enzymes from similar pathways to generate compounds in non-native hosts [126].

1.10 Objectives of this study

Addressing the challenges of natural product production, we propose an overall objective to develop a strategy to discover the genetic underpinnings of orphan compounds and use this information to engineer strains that can overproduce their intermediates as well as the compounds themselves.

This overarching objective is broken down into four sub-objectives. First to identify and catalog the clusters and their genes that are responsible for the production of secondary metabolites in the genome of *A. niger* NRRL3. Secondly, to overexpress transcription factor genes predicted to regulate SM clusters (clustered transcription factors) within *A. niger* NRRL3 in an effort to generate their secondary metabolites. Thirdly, to reconstruct biosynthetic pathways for compounds for which the genetic underpinnings are currently unknown. Lastly, to generate deletion mutants based on reconstructed pathways to accumulate intermediate compounds and test the robustness of our pathway reconstruction method by predicting the metabolic outcomes of the deleted genes.

Materials and Methods

2.1 Bioinformatics methods

2.1.1 Backbone enzyme identification

All backbone enzymes were initially identified via BLASTP searches using previously characterized backbone enzyme sequences from other Aspergilli as queries (E-value ≥ 1x 10^{-5}). All hits were evaluated for acceptance into their respective categories by protein domain content using the online applications Pfam (http://pfam.xfam.org/), conserved domain database (CDD) (http://www.ncbi.nlm.nih.gov/cdd/), and InterProScan 5 (http://www.ebi.ac.uk/interpro/search/sequence-search). Domain annotations for each enzyme were accepted when at least two databases agreed on annotation. Sequences were accepted as PKS, NRPS and HPN enzymes when at least all respective minimal domains were present with the exception of PKS NRRL3_01804 whose ACP domain does not have a complete GX(H/D)S(L/I) (G→P) amino acid motif but which still may be functional based on previous work with non-canonical ACP domains [29, 127]. The PKS-like and NRPS-like enzymes were accepted when at least two databases agreed on at least two domains which are contiguous for specific enzyme categories and at least one was a minimal domain. For domains which were not documented in the domain databases, multiple sequence alignments were performed to confirm the presence of previously published amino acid motifs and conserved residues. This process was limited to start unit acyl transferase [128] (SAT), product template [129] (PT) and FAS ACP domains [80].

An additional round of BLASTP searches was subsequently carried out using isolated domain pairs (from enzymes in the first round) which appear contiguously within enzyme categories and contain at least one minimal domain (E-value cut-off of 1x10^{5}). This was done to find more backbone enzymes by reducing gap penalties from BLAST alignments. I surmised that using domain pairs that are always contiguous in each backbone class would have fewer and shorter gaps when used as a query sequence compared to using the full backbone enzyme incurring small gap penalties. For instance, a BLAST search using a full non-reducing PKS against the genome would incur a gap penalty when aligned with a highly reducing PKS due to the intervening reducing domains between the AT and ACP domains in the highly reducing PKS.

Even a highly reducing PKS may incur gap penalties when aligned with another such PKS from uncommon domains like MT or alcohol dehydrogenase domains. This process was particularly helpful with NRPS enzymes given the higher variability in their domain organization than PKS enzymes.

For FAS enzymes, given the lack of variation in their domain organization compared to PKS enzymes, the FAS alpha and beta subunit enzymes were accepted into their respective categories if the prospective enzyme contained the general domain architecture (the PPT domain notwithstanding) of the fungal FAS alpha or beta domains as outlined by Jenni, S. et al [80]. All DMAT, sesquiterpene cyclase, and type III PKS single domain enzymes were accepted based on InterProScan families (IPR017795, IPR024652 and IPR011141 respectively) and CDD domains (PT-DMATS_CynD, Isoprenoid_Biosyn_C1 and BH0617 respectively).

2.1.2 Backbone enzyme comparison between *A. niger* NRRL3 and *A. niger* CBS 513.88

To compare closely related strains in terms of secondary metabolite genes, protein sequences of identified enzyme backbone classes from *A. niger* NRRL3 as well as *A. niger* CBS 513.88 were used as queries against each other's genomes to identify counterparts and unique enzymes between the strains. Annotated enzymes from *A. niger* CBS 513.88 were obtained from the AspGD website (http://www.aspergillusgenome.org/) Backbone enzymes with %ID $\geq$ 85 (E-value $\geq$ 1x 10^{-5}) from BLASTP queries were designated counterparts and all others were considered unique to either strain.

2.1.3 Backbone enzyme remnant and remnant cluster analysis

In order to determine if backbone-like enzymes had complete counterparts in other species, PKS-like and NRPS-like enzyme sequences as well as all possible singlet domain types in the multi-domain secondary metabolite backbone enzymes were used as queries against the fungal genomes available at the AspGD website (%ID $\geq$ 50, E-value $\geq$ 1x 10^{-5})). Any backbone-like enzyme which had a full counterpart in another species was labelled as a remnant.

We next examined the homology of other genes around identified remnant backbone genes in species with complete counterparts to identify missing, if any, SM cluster genes.

Homology was visualized using the Sybil online application at the AspGD website (http://aspgd.broadinstitute.org/cgi-bin/asp2_v3/shared/index.cgi?site=asp2_v9). Since the Sybil application could not find homologues of the remnant enzymes (possibly due to their highly truncated structure), accessory enzymes in their vicinity were used as anchor points for Sybil homology searches.

2.1.4 Secondary metabolite cluster definition

Secondary metabolite clusters were defined using the clusters which were previously annotated in *A. niger* CBS 513.88 as a reference [27]. These clusters were modified to incorporate our updated gene annotation data and took precedence over the original annotations including those defined by Sybil synteny alignments. Apart from automated backbone discovery (by SMURF and antiSMASH) and syntenic cluster definition, all methods for cluster definition in this study followed those outlined for CBS 513.88 [27] for those clusters whose backbone enzymes were unique to this study.

2.1.5 Assignment of secondary metabolites to annotated SM clusters and mitochondrial peptide localization sequences

Metabolic products were assigned to specific PKS, NRPS and HPN clusters based on experimentally defined orthologous clusters or backbones. Experimentally defined backbone enzymes with known clusters and metabolic products were used as a query sequence for BLASTP against the NRRL3 genome. The experimentally defined backbone enzymes which had a match in the NRRL3 genome (%ID $\geq$ 40%, query coverage $\geq$ 80%) were set aside and their accompanying accessory enzymes were then queried against the NRRL3 genome. Backbone and accessory enzymes from literature were assigned an *A. niger* NRRL3 counterpart where % sequence identity was $\geq$ 80%, 50% and 40% (E-value $\geq$ 1x 10^{-5}) for other *A. niger* strains, other species of *Aspergillus*, and non-*Aspergillus* filamentous fungi respectively (query coverage $\geq$ 80%). Where BLASTP top hits from other genomes were co-localized in *A. niger* NRRL3, the group was accepted as an orthologous cluster. Lastly, to determine subcellular localization sequences for mitochondrial proteins, we used the targetP 1.1 online application

(http://www.cbs.dtu.dk/services/TargetP/) to look for mitochondrial targeting and processing peptides in protein sequences [55].

2.2 Methods for alkyl citric acid cluster investigation

2.2.1 Strains and culture conditions

Aspergillus niger PY11 (cspA− pyrG− ΔGla::hiG), a derivative of NRRL3, was used for all fungal transformations and overexpression studies. The culture conditions and transformation protocol used for this study have been described previously [130]. The DH5α strain of *E. coli* was used for propagation of cloned plasmids. For metabolite production, conidia ($2x10^6$ conidia/mL) from *A. niger* transformants were inoculated into liquid MM "J" medium [130] and incubated without shaking at 30°C for 5 days. For transcriptome analysis, conidia at $2x10^6$/mL were inoculated in minimal medium [131] containing 2% fructose, 0.1% yeast extract and 0.1% casamino acids. Cultures were incubated at 30°C and with shaking at 220 rpm. After 24 hours of growth, mycelia were collected by gravity filtration using Miracloth (EMD Millipore – 475855, Darmstadt, Germany) through a Buchner funnel. Following washing with water, approximately one teaspoon (~5 mL) of mycelia was transferred to 50 mL of minimal medium containing 2% maltose. Shaking cultures containing the mycelia were incubated at 30°C and at 220 rpm for 2 hours, harvested by filtration as above. The harvested mycelia were then frozen in liquid nitrogen for RNA extraction.

2.2.2 Construction of overexpression vectors

Two plasmid vectors were constructed to overexpress the regulator genes within our defined clusters. The difference between them is the marker gene used for selection of *A. niger* transformants (Figure 10). In the first vector, the plasmid ANIp7 [132] was amplified by PCR using primers containing a 21-nucleotide adapter sequence (Figure 10, Table 3). The plasmid insert transcription factor was prepared by PCR from genomic DNA obtained using a GeneJET Genomic DNA Purification Kit (Thermo K0721, Thermo Scientific, Grand Island, NY USA). Adapter sequences from insert and vector amplification primers were designed to be complementary allowing for annealing for ligation-independent cloning [133]. This vector was used for the initial screening to identify SM regulators and for transcriptome analysis. In the

second expression vector, the *pyrG* selection marker was replaced by an *amdS* selection marker (Figure 10, Table 3) and was used for overexpression of the NRRL3_11765 transcription factor to construct a strain which would subsequently be used for gene deletions. Gene deletions were then selected by using the *pyrG* marker prior to screening.

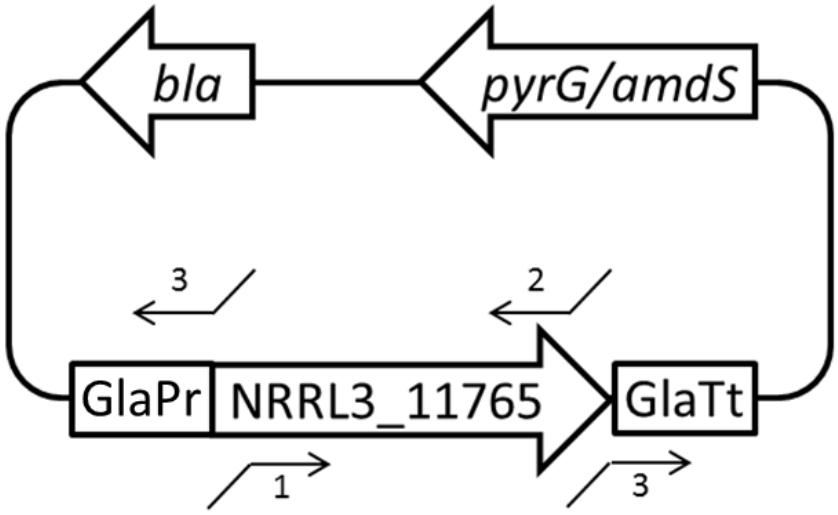


Figure 10. Overexpression cassette containing the NRRL3_11765 transcription factor. The cassettes contain the bacterial ampicillin selection marker (*bla*), a promoter/terminator (GlaPr, GlaTt) and the fungal selection markers *pyrG* for the initial screen and *amdS* for expression in deletion strains. Numbered bent arrows indicated the primers (with the adapter sequences at the 5' end) used for construction of the plasmid and insert fragments; refer to numbers in Table 3 for primer sequences.

Table 3. The list of primers used for plasmid and gene deletion cassette assembly. Underlined sections represent adapter sequences.

No.	Primer Name	Sequence 5' --> 3'
1	NRRL3_11765 Fw	TACTTCCAATCCAATCCATTTGACGATATGGCTACTCCCCGCGCATC
2	NRRL3_11765 Re	TTATCCACTTCCAATCCATTTGTCACTGCATCAATCCCAACGCAG
3	ANIp7 Fw	AAATGGATTGGAAGTGGATAACTTAATTAAGTTTAAACG
4	ANIp7 Re	AAATGGATTGGATTGGAAGTACTGATCTAGCGTGTAATG
5	5' NRRL3_11766 Fw	AATTCCCAGCTCCAAGTGC
6	5' NRRL3_11766 Re	AAATGGATTGGAAGTGGATAACGGAGACTTCTGATGAATATGTGG
7	3' NRRL3_11766 Fw	AAATGGATTGGATTGGAAGTACGCCAATCTTGACCAGGATG
8	3' NRRL3_11766 Re	ACATTTGGTCGGGATAACAACAAC
9	*pyrG* sel. marker Fw	TTATCCACTTCCAATCCATTTGCCCCTTTTAGTCAATACCG
10	*pyrG* sel. marker Re	TACTTCCAATCCAATCCATTTGCGCAACTTCCTCGAGAAC
11	NRRL3_11766 ext Fw	ATTCTCACCTATGCAGGCAC
12	NRRL3_11766 ext Re	CCATGCTCGAGGGTAAAAGC
13	5' NRRL3_03750 Fw	AAATCCGCATGAAGAGGACC
14	5' NRRL3_03750 Re	AAATGGATTGGAAGTGGATAACTCCAGGACAAACAAGAGGCAC
15	3' NRRL3_03750 Fw	AAATGGATTGGATTGGAAGTACCGTGGAGTATTTGTTCGGAATG
16	3' NRRL3_03750 Re	AGTTAGACTGGTATTCATTAAGGAGAG
17	NRRL3_03750 ext Fw	ACTTACCATTGCTTTAGTGATCC
18	NRRL3_03750 ext Re	GACTTTCCTGACTTGCAACTC
19	5' NRRL3_08383 Fw	GGGATTCAAGGGGTAAATGGAAC
20	5' NRRL3_08383 Re	AAATGGATTGGAAGTGGATAACCCCTCATAGCACGACGAAATC
21	3' NRRL3_08383 Fw	AAATGGATTGGATTGGAAGTACGAGGAATGGAATCTACTTTTGGTG
22	3' NRRL3_08383 Re	TTGTTGATTTCCGGCGTAAGC
23	NRRL3_08383 ext Fw	CTGGTTGGATCGTTGTTTTG
24	NRRL3_08383 ext Re	TCTCATTCCTCATCCCCTC

2.2.3 Gene deletion cassette construction and verification in *A. niger* NRRL3_11765 overproducing strain

Three linear cassettes were designed to delete a 2-methylcitrate dehydratase gene (NRRL3_11766), a monooxygenase gene (NRRL3_03750), and an o-methyltransferase gene (NRRL3_08383) thought to be involved in NRRL3_11765 upregulated pathway (Figure 11-13). The cassettes consisted of two homologous DNA sequences flanking each target gene and fused with an intervening *pyrG* selection marker by overlap PCR. The overlap PCR used the same primer adapter sequence as described in section 2.2.2 (Table 3). Genomic sequence information

used in this study was obtained from the *A. niger* genome resource at www.fungalgenomics.ca. Deletion mutants were screened by PCR using primers binding upstream of the 5' flank and downstream of the 3' flank (Figure 11-13).

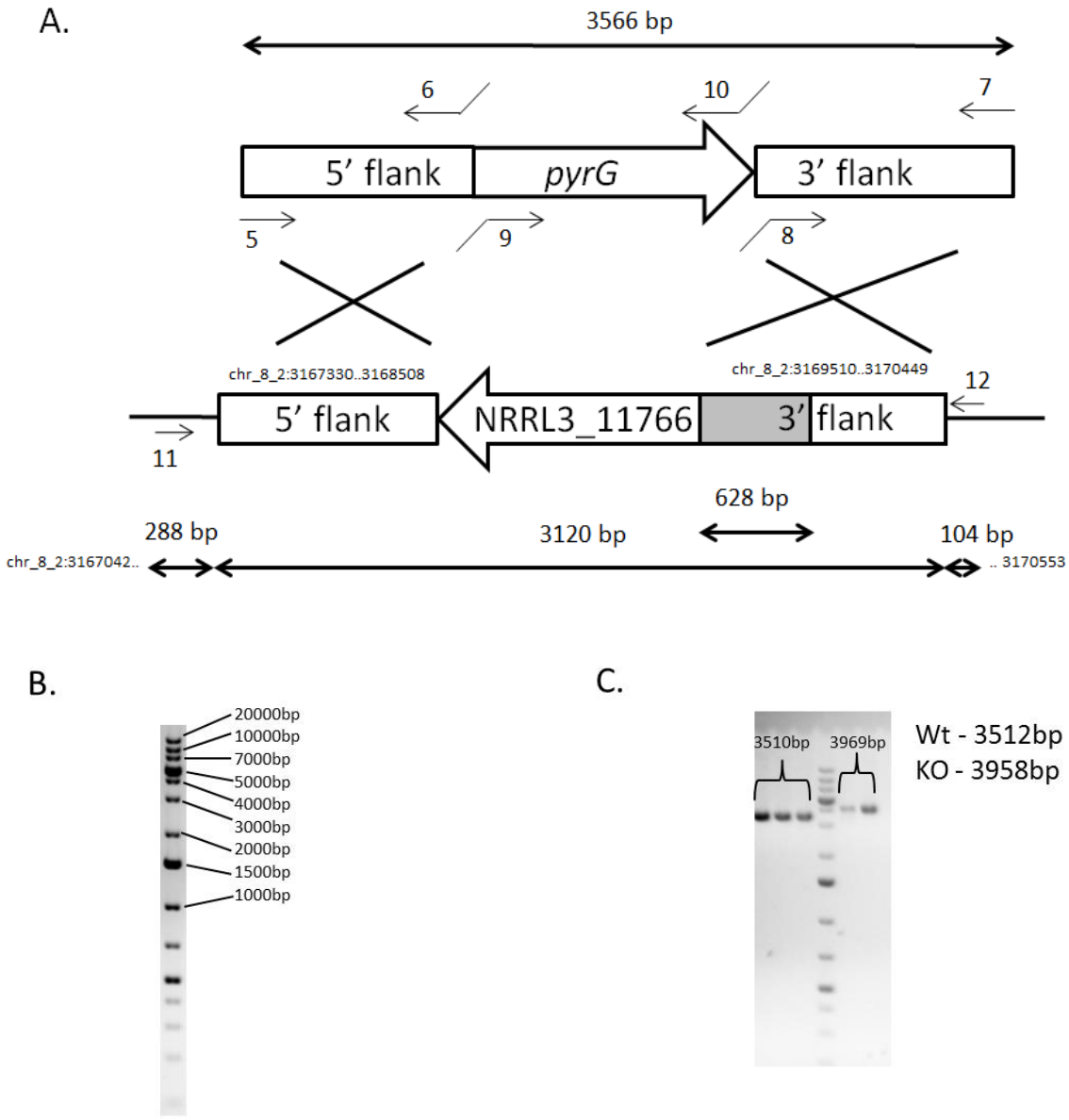


Figure 11. (A) Deletion cassette for NRRL3_11766. Crossed lines indicate regions of homologous recombination between the cassette sequence (top) and the target gene sequence (bottom). Genome coordinates of 5' and 3' flanks are shown above each flank. A gray shaded box in the target gene sequence indicates the area of overlap between the 3' flank and NRRL3_11766. Numbered arrows within cassettes indicate the primers used for PCR amplification of DNA fragments for their assembly (refer to numbers in Table 3 for primer sequences). Numbered arrows outside of 5' and 3' flanks indicate primers used for PCR verification of homologous recombination in transformants. Bent arrows indicate primers with adapter sequences. (B) Gene ruler 1kb plus. (C) Result of PCR product screen for the NRRL3_11766 gene deletion (KO) and wildtype (wt) strains using primers 11 and 12.

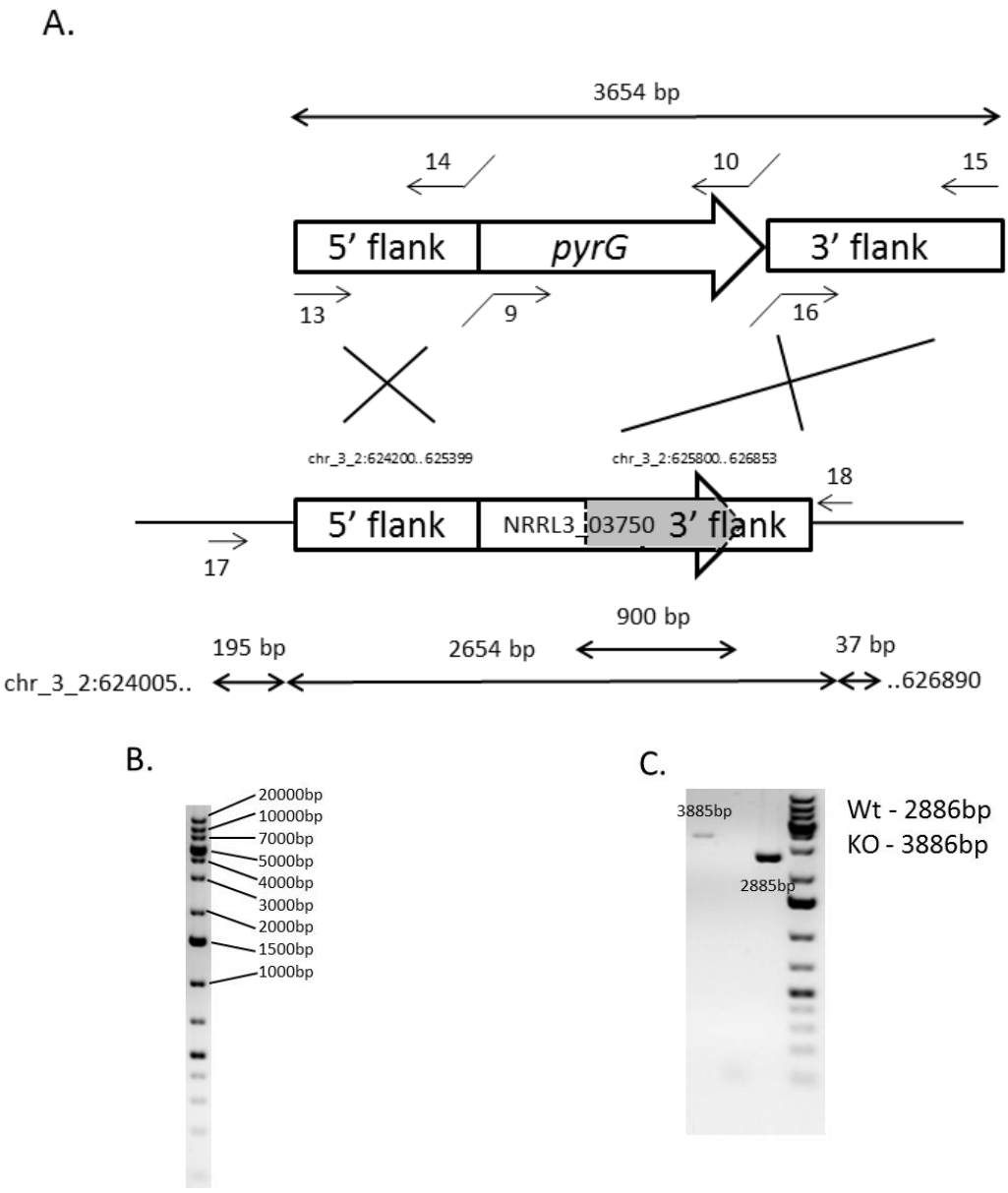


Figure 12 (A) Deletion cassette for NRRL3_03750. Crossed lines indicate regions of homologous recombination between the cassette sequence (top) and the target gene sequence (bottom). Genome coordinates of 5' and 3' flanks are shown above each flank. A gray shaded box in the target gene sequence indicates the area of overlap between the 3' flank and NRRL3_03750. Numbered arrows within cassettes indicate the primers used for PCR amplification of DNA fragments for their assembly (refer to numbers in Table 3 for primer sequences). Numbered arrows outside of 5' and 3' flanks indicate primers used for PCR verification of homologous recombination in transformants. Bent arrows indicate primers with adapter sequences. (B) Gene ruler 1kb plus. (C) Result of PCR product screen for the NRRL3_03750 gene deletion (KO) and wildtype (wt) strains using primers 17 and 18.

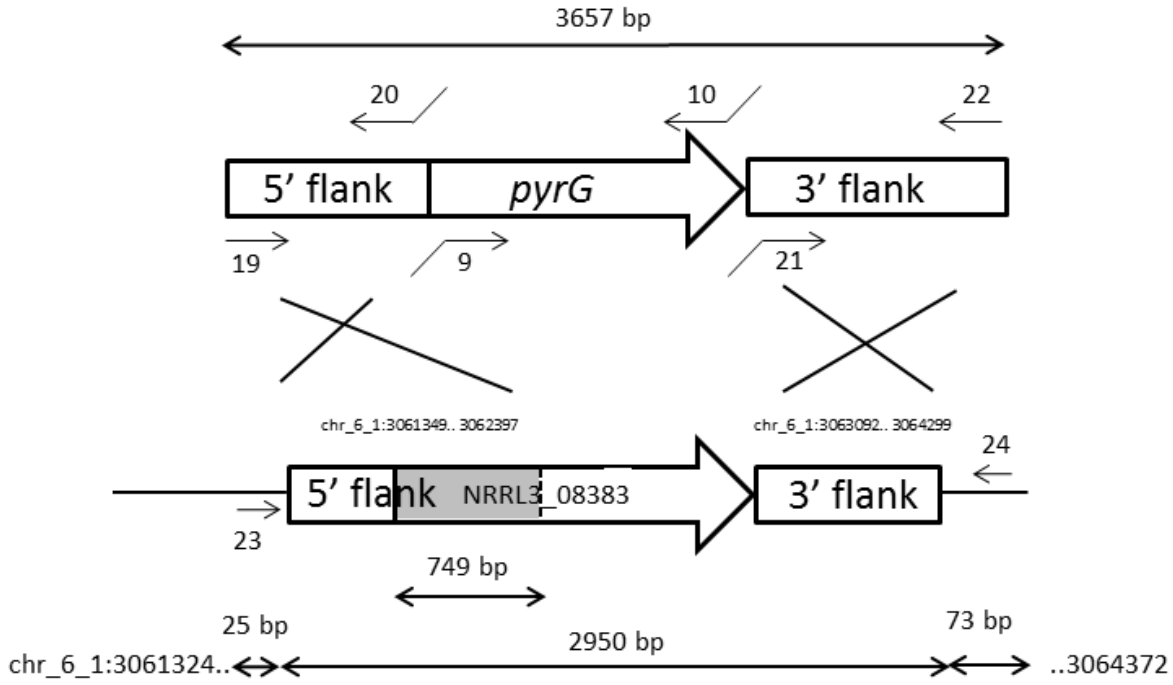


Figure 13. Deletion cassette for NRRL3_08383. Crossed lines indicate regions of homologous recombination between the cassette sequence (top) and the target gene sequence (bottom). Genome coordinates of 5' and 3' flanks are shown above each flank. A gray shaded box in the target gene sequence indicates the area of overlap between the 3' flank and NRRL3_08383. Numbered arrows within cassettes indicate the primers used for PCR amplification of DNA fragments for their assembly (refer to numbers in Table 3 for primer sequences). Numbered arrows outside of 5' and 3' flanks indicate primers which were used for PCR verification of homologous recombination in transformants. Bent arrows indicate primers with adapter sequences.

2.2.4 Transcriptome sequencing and analysis

Total RNA was extracted from frozen (by liquid nitrogen) mycelia using TRIzol reagent (Thermo Scientific, Grand Island, NY USA) as described [134]. A cDNA library was constructed using an Illumina TruSeq Stranded mRNA Sample Preparation Kit and sequenced to generate 100-bp reads on an Ilumina HiSeq 2000 system at the McGill University Génome Québec Innovation Centre. Differential gene expression analysis was carried out using the DEseq Bioconductor software package [135]. Genes were considered significantly up or down regulated due to NRRL3_11765 overexpression when p-value < $2x10^{-4}$. Genes expressed at very

low levels, below 10 fragments per kilobase of transcript per million (FPKM), in both the control and overexpression strains were removed from analysis.

2.2.5 Preparation of metabolites for mass spectrometric analysis

To prepare extracellular metabolites in growth media for mass spectrometry, mycelia were removed following cultivation by centrifugation at 16,000 x g for 45 minutes. The supernates were transferred to fresh tubes and mixed with an equal volume of 99% purity grade cold methanol (CH_3OH) (-20°C). After resting on ice for 10 minutes, the samples were centrifuged at 16,000 x g for 45 minutes to remove precipitated proteins. The resulting supernates were transferred to fresh tubes and mixed with an equal volume of 0.1% formic acid.

For the preparation of intracellular metabolites, fungal mycelia were ground with liquid nitrogen by mortar and pestle. The ground powder was dissolved in water and sonicated while on ice for 5 minutes, with alternate cycles consisting of 5 seconds off and 15 seconds on. After sonication, 1 mL of the samples were centrifuged at 16,000 x g for 45 minutes and the supernates were prepared for mass spectrometry analysis as described for extracellular metabolites and transferred to amber vials with inserts for HPLC and mass spectrometry (Wheaton Co, Product 09-2200-101, Millville, NJ USA).

2.2.6 Analysis of metabolites by mass spectrometry

Electrospray LC-MS analyses were performed on a 7-Tesla Finnigan LTQ-FT-ICR mass spectrometer (Thermo Electron Corporation, San Jose, CA). Ionization voltage used was 4900 V in positive mode and 3700 V in negative mode. Scan range was from 100 to 1400 m/z at a 50000-resolution setting. The solvent delivery system used was a Series 200 auto sampler and micropump (Perkin Elmer, Waltham, MA). Injection volume was 10 µL and flow rate was 250 µL/minute. Reversed-phase liquid chromatography (RPLC) separation was performed using an Eclipse C18 3.5 µm, 2.1 x 150 mm column (Agilent, Santa Clara, CA). The solvents used to generate the gradient during reversed-phase liquid chromatography (RPLC) separation were 0.1 % formic acid in water for Solvent A and 0.1% formic acid in acetonitrile for Solvent B. The gradient was used to elute the metabolites: 5% B isocratic for 1 minute, increased to 95% B in 10

minutes, isocratic at 95% for 1 minute, decreased to 5% B in 0.1 minute, and isocratic at 5% for 5.9 minutes.

Total ion chromatograms, with a range of 190m/z – 300m/z, were generated using the XcaliburTM software version 2.2 (Thermo Scientific 2011). Aligned peak identification and extracted ion chromatograms (EICs) construction were carried out using Maven (2011 version 6.12) scanning at a minimum peak intensity threshold of $5x10^4$ and a mass tolerance of 5 ppm [136]. Peak identification was carried out primarily on positive mode mass spectra for consistency using a manually assembled database ("secondary metabolite database") of secondary metabolites from literature searches. Negative mode EICs were used instead where an identified compound could not be detected above threshold in positive mode.

2.2.7 Analysis by nuclear magnetic resonance (NMR)

Secondary metabolites from *A. niger* overexpressing transcription factor NRRL3_11765 (*chaR*^{OE}) were extracted from 700 ml of culture with 1L of ethyl acetate (EtOAc) to remove sugars and other solutes from growth media. The extract was dried *in vacuo* to yield an oily material (6g) which was dissolved in 1 mL of chloroform ($CHCl_3$), and then applied to a silica gel column (57 × 508 mm, 250g, Zeoprep 60, 40 - 60 mm). Elution factions included 100:0, 100:1, 75:1, 50:1, 25:1, 1:1, and 0:100 (v/v) of $CHCl_3$- CH_3OH solvents, 250ml each. The 2-carboxymethyl-3-hexylmaleic acid (5.3 g) compound was eluted in the two first fractions (100:0, 100:1), while hexylitaconic acid (150 mg) and the hydroxylated form of hexylitaconic acid (200 mg) were collected from fractions 75:1 and 25:1, respectively. Hexylcitraconic acid (100 mg) and the carbonylated form of hexylitaconic acid (6 mg) were eluted at fraction 50:1 in different collection tubes. All these SMs were dissolved in $CDCl_3$ (Sigma-Aldrich) for structural investigation by NMR, except the hydroxylated hexylitaconic acid (a polar molecule), which was dissolved in CD_3OD (Sigma-Aldrich) for later NMR investigation.

Twenty millilitres of culture filtrate from the *chaR*^{OE} strain containing a 2-methylcitrate dehydratase deletion (NRRL3_11766) were extracted twice with 40 mL of EtOAc. The extract was dried *in vacuo* to a brown powder. Twenty four milligrams of the powder were dissolved in 3 mL of 1:1 mixture of acetonitrile containing 0.1% trifluoroacetic acid (TFA) and 0.1% TFA in high-pressure liquid chromatography (HPLC)-grade water which were later used as HPLC

solvent mobile phase A and B, respectively. The resulting preparation was injected (95 µL/injection) into an Eclipse-C18 column (4.6 × 150 mm) and the flow rate was set at 1 mL/minute. The HPLC gradient included a linear increase from 10% to 95% of mobile phase B in 10 minutes, and a constant condition of 95% B in 2 minutes. Mobile phase B was returned to 10% in 1 minute and then held for 6 minutes to allow the column re-equilibration process. The detector was set at a wavelength of 210 nm. Under this condition, hexylcitric acid and its anhydride were co-eluted at 6.9 minute. The collected fraction, from 6.6 minute to 7.1 minute, was extracted with EtOAc and concentrated to yield 20 mg of yellow syrup. For NMR analysis the syrup was dissolved in dimethyl sulphoxide-d6. All NMR spectra were recorded with a Varian VNMRS-500 MHz (Department of Chemistry and Biochemistry, Concordia University, Montreal, Quebec, Canada) at 25°C.

Results

3.1 Secondary metabolite genes and gene clusters in *A. niger* NRRL3 genome.

3.1.1 Secondary metabolite backbone genes and clusters defined for *A. niger* NRRL3.

Bioinformatics analysis of the *A. niger* NRRL3 genome revealed a total of 96 secondary metabolite backbone genes assembled into 84 gene clusters (Table 4).

Table 4. Secondary metabolite clusters of *A. niger* NRRL3 and *A. niger* CBS 513.88.

Secondary metabolite gene cluster	NRRL 3	CBS 513.88	Description	Domain(s) of backbone enzymes
Azanigerone cluster	NRRL3_00140	An09g01790	esterase/lipase AzaC	
Azanigerone cluster	NRRL3_00141	An09g01800	acyltransferase AzaD	
Azanigerone cluster	NRRL3_00142	An09g01810	ketoreductase AzaE	
Azanigerone cluster	NRRL3_00143	An09g01820	AMP-dependent CoA ligase AzaF	
Azanigerone cluster	NRRL3_00144	An09g01830	FAD-dependent oxygenase AzaG	
Azanigerone cluster	NRRL3_00145[a]	An09g01840	salicylate monooxygenase AzaH	
Azanigerone cluster	NRRL3_00146	An09g01850	cytochrome P450 AzaI	
Azanigerone cluster	NRRL3_00147	An09g01860	polyketide synthase AzaA	SAT-KS-AT-ACP-MT-KR
Azanigerone cluster	NRRL3_00148[a]	An09g01870	fungal-specific transcription factor AzaR	
Azanigerone cluster	NRRL3_00149	An09g01880	zinc-type alcohol dehydrogenase AzaJ	
Azanigerone cluster	NRRL3_00150	An09g01900	haem-degrading HbpS-like protein	
Azanigerone cluster	NRRL3_00151	An09g01910	MFS-type efflux transporter AzaK	
Azanigerone cluster	NRRL3_00152	An09g01920	FAD-dependent oxygenase AzaL	
Azanigerone cluster	NRRL3_00153[a]	An09g01930	polyketide synthase AzaB	KS-AT-DH-MT-KR-ER-KR-ACP
Carlosic acid and agglomerin cluster	NRRL3_11025	An08g03730	oxoglutarate/iron-dependent dioxygenase CaaD	
Carlosic acid and agglomerin cluster	NRRL3_11026	An08g03740	trans enoylreductase CaaB	
Carlosic acid and agglomerin cluster	NRRL3_11027	An08g03750	major facilitator superfamily protein	
Carlosic acid and agglomerin cluster	NRRL3_11028	An08g03760	SnoaL-like domain-containing protein	
Carlosic acid and agglomerin cluster	NRRL3_11029[a]	An08g03770	fungal-specific transcription factor CaaR	
Carlosic acid and agglomerin cluster	NRRL3_11030	An08g03780	cytochrome P450 CaaC	
Carlosic acid and agglomerin cluster	NRRL3_11031	An08g03790	hybrid polyketide synthase/non-ribosomal peptide synthetase CaaA	KS-AT-DH-MT-KR-ACP--C-Rpt-A-(T)-(Te)
Conidial pigmentation gene	NRRL3_00462[a]	An09g05730	polyketide synthase AlbA/FwnA	SAT-KS-AT-PT-ACP-ACP-Te
Conidial pigmentation gene	NRRL3_01039[a]	An14g05350	hydrolase AygA/OlvA	
Conidial pigmentation gene	NRRL3_01040[a]	An14g05370	multicopper oxidase BrnA	
Conidial pigmentation gene	NRRL3_02783[a]	An01g14010	multicopper oxidase Mco1A	

Fumonisin cluster	NRRL3_02178	An01g06820	bifunctional P-450:NADPH-P450 reductase Fum6	
Fumonisin cluster	NRRL3_02179	An01g06830	short-chain dehydrogenase/reductase family protein	
Fumonisin cluster	NRRL3_02180	An01g06840	AMP-dependent synthetase/ligase Fum10	
Fumonisin cluster	NRRL3_02181	An01g06850	iron-containing alcohol dehydrogenase Fum7	
Fumonisin cluster	NRRL3_02182	An01g06860	phytanoyl-CoA dioxygenase Fum9	
Fumonisin cluster	NRRL3_02183	An01g06870	aminotransferase, class I/class II Fum8	
Fumonisin cluster	NRRL3_02184	An01g06880	NAD-dependent epimerase/dehydratase family protein Fum13	
Fumonisin cluster	NRRL3_02185	An01g06890	non-ribosomal peptide synthetase-like protein Fum14	T-C
Fumonisin cluster	NRRL3_02186	An01g06900	fungal-specific transcription factor Fum21	
Fumonisin cluster	NRRL3_02187	An01g06910	cytochrome P450 Fum15	
Fumonisin cluster	NRRL3_02188	An01g06920	ABC transporter Fum19	
Fumonisin cluster	NRRL3_02189	An01g06930	polyketide synthase Fum1	KS-AT-DH-MT-ER-KR-ACP
Fumonisin cluster	NRRL3_02190	An01g06940	major facilitator superfamily protein	
Kotanin cluster	NRRL3_07482	An04g09570	major facilitator superfamily protein	
Kotanin cluster	NRRL3_07483	An04g09560	cell wall mannoprotein 1-like protein	
Kotanin cluster	NRRL3_07484	An04g09550	flavin monooxygenase-like protein	
Kotanin cluster	NRRL3_07485[a]	An04g09540	cytochrome P450 KtnC	
Kotanin cluster	NRRL3_07486[a]	An04g09530	polyketide synthase KtnS	SAT-KS-AT-PT-ACP
Kotanin cluster	NRRL3_07487[a]	An04g09520	O-methyltransferase KtnB	
NRRL3_00030 cluster	NRRL3_00025	An09g00390	FAD-dependent oxidoreductase domain-containing protein	
NRRL3_00030 cluster	NRRL3_00026	An09g00400	amino acid/polyamine transporter I family protein	
NRRL3_00030 cluster	NRRL3_00027	An09g00420	HAD-like domain-containing protein	
NRRL3_00030 cluster	NRRL3_00028	An09g00430	fungal-specific transcription factor	
NRRL3_00030 cluster	NRRL3_00029	An09g00440	hypothetical protein	
NRRL3_00030 cluster	NRRL3_00030	An09g00450	non-ribosomal peptide synthetase-like protein	A-T-NADB
NRRL3_00036 cluster	NRRL3_00036	An09g00520	non-ribosomal peptide synthetase	T-C-A-T-C-Rpt
NRRL3_00036 cluster	NRRL3_00037	An09g00530	FAD-binding domain-containing protein	
NRRL3_00036 cluster	NRRL3_00038	An09g00540	hypothetical protein	
NRRL3_00036 cluster	NRRL3_00039	An09g00550	major facilitator superfamily protein	
NRRL3_00036 cluster	NRRL3_00040	An09g00560	cytochrome P450 family protein	
NRRL3_00036 cluster	NRRL3_00041	An09g00570	NAD(P)-binding domain-containing protein	
NRRL3_00036 cluster	NRRL3_00042	An09g00580	fungal-specific transcription factor	
NRRL3_00036 cluster	NRRL3_00043	An09g00590	FAD-binding domain-containing protein	

NRRL3_00036 cluster	NRRL3_00044	An09g00600	NmrA-like family protein	
NRRL3_00036 cluster	NRRL3_00046	NFG	hypothetical protein	
NRRL3_00036 cluster	NRRL3_00048	An09g00620	short-chain dehydrogenase/reductase family protein	
NRRL3_00102 cluster	NRRL3_00099	An09g01260	fatty acyl-CoA reductase-like protein	
NRRL3_00102 cluster	NRRL3_00100	An09g01270	cytochrome P450	
NRRL3_00102 cluster	NRRL3_00101	An09g01280	cell wall mannoprotein 1-like protein	
NRRL3_00102 cluster	NRRL3_00102	An09g01290	polyketide synthase	KS-AT-DH-MT-ER-KR-ACP
NRRL3_00102 cluster	NRRL3_00103	An09g01300	organic solute transporter subunit alpha/transmembrane protein 184 family protein	
NRRL3_00102 cluster	NRRL3_00104	An09g01310	fungal-specific transcription factor	
NRRL3_00102 cluster	NRRL3_00105	An09g01320	FAD-binding domain-containing protein	
NRRL3_00102 cluster	NRRL3_00106	An09g01330	squalene cyclase family protein	
NRRL3_00102 cluster	NRRL3_00107	An09g01340	cytochrome P450 family protein	
NRRL3_00135/NRRL3_00138/NRRL3_00159 cluster	NRRL3_00134	An09g01680	RTA-like protein	
NRRL3_00135/NRRL3_00138/NRRL3_00159 cluster	NRRL3_00135	An09g01690	non-ribosomal peptide synthetase	A-T-C-C-A-T-C
NRRL3_00135/NRRL3_00138/NRRL3_00159 cluster	NRRL3_00136	An09g01700	ABC transporter type 1, transmembrane domain-containing protein	
NRRL3_00135/NRRL3_00138/NRRL3_00159 cluster	NRRL3_00137	An09g01710	alpha/beta hydrolase fold-1 domain-containing protein	
NRRL3_00135/NRRL3_00138/NRRL3_00159 cluster	NRRL3_00138	An09g01740	fatty acid synthase beta subunit	AT-ER-DH-MPT
NRRL3_00135/NRRL3_00138/NRRL3_00159 cluster	NRRL3_00154	An09g01940	RTA-like protein	
NRRL3_00135/NRRL3_00138/NRRL3_00159 cluster	NRRL3_00139	An09g01760	HotDog domain-containing protein	
NRRL3_00135/NRRL3_00138/NRRL3_00159 cluster	NRRL3_00155	An09g01950	cytochrome P450	
NRRL3_00135/NRRL3_00138/NRRL3_00159 cluster	NRRL3_00156	An09g01970	oxidoreductase domain-containing protein	
NRRL3_00135/NRRL3_00138/NRRL3_00159 cluster	NRRL3_00157	An09g01990	branched-chain amino acid aminotransferase-like protein	
NRRL3_00135/NRRL3_00138/NRRL3_00159 cluster	NRRL3_00158	An09g02000	cytochrome P450 family protein	
NRRL3_00135/NRRL3_00138/NRRL3_00159 cluster	NRRL3_00159	An09g02010	fatty acid synthase alpha subunit	MPT-ACP-KR-KS
NRRL3_00166 cluster	NRRL3_00166	An09g02100	polyketide synthase-like protein	[KR-KS-AT-DH-ACP]-ACP-MT-Te
NRRL3_00166 cluster	NRRL3_00167	An09g02120	isochorismatase family protein	
NRRL3_00166 cluster	NRRL3_00168	An09g02150	FAD-binding domain-containing protein	
NRRL3_00204 cluster	NRRL3_00202	An09g02580	major facilitator superfamily protein	
NRRL3_00204 cluster	NRRL3_00203	An09g02590	cytochrome P450	
NRRL3_00204 cluster	NRRL3_00204	An09g02610	sesquiterpene cyclase	Isoprenoid_Biosyn_C1

NRRL3_00204 cluster	NRRL3_00205	An09g02620	fungal-specific transcription factor	
NRRL3_00410 cluster	NRRL3_00405	An09g05050	gurmarin/antimicrobial peptide-like protein	
NRRL3_00410 cluster	NRRL3_00406	An09g05060	fungal-specific transcription factor	
NRRL3_00410 cluster	NRRL3_00407	An09g05070	major facilitator superfamily protein	
NRRL3_00410 cluster	NRRL3_00408	An09g05080	aminotransferase class I and II family protein	
NRRL3_00410 cluster	NRRL3_00409	An09g05100	impact family protein	
NRRL3_00410 cluster	NRRL3_00410	An09g05110	non-ribosomal peptide synthetase-like protein	A-T-Te
NRRL3_00410 cluster	NRRL3_00411	An09g05120	carboxylesterase family protein	
NRRL3_00410 cluster	NRRL3_00412	An09g05130	catechol oxidase	
NRRL3_00410 cluster	NRRL3_00413	An09g05140	NmrA-like family protein	
NRRL3_00410 cluster	NRRL3_00414	An09g05150	intradiol ring-cleavage dioxygenase core domain-containing protein	
NRRL3_00430 cluster	NRRL3_00430	An09g05340	polyketide synthase	KS-AT-DH-ER-KR-ACP
NRRL3_00430 cluster	NRRL3_00431	An09g05350	tannase/feruloyl esterase family protein	
NRRL3_00755 cluster	NRRL3_00755	An14g01910	hybrid polyketide synthase/non-ribosomal peptide synthetase	KS-AT-DH-MT-KR-ACP--C-Rpt-A-(T)
NRRL3_00755 cluster	NRRL3_00756	An14g01940	zinc-type alcohol dehydrogenase superfamily protein	
NRRL3_00755 cluster	NRRL3_00757	An14g01950	alcohol acetyltransferase/N-acetyltransferase family protein	
NRRL3_00755 cluster	NRRL3_00758	An14g01960	anoctamin-like protein	
NRRL3_00998 cluster	NRRL3_00997	An14g04840	methyltransferase VrtF-like protein	
NRRL3_00998 cluster	NRRL3_00998	An14g04850	hybrid polyketide synthase/non-ribosomal peptide synthetase	A-T--KS-AT-KR-ACP
NRRL3_01334 cluster	NRRL3_01332	An13g03060	CDR ABC transporter domain-containing protein	
NRRL3_01334 cluster	NRRL3_01333.1	An13g03050	hypothetical protein	
NRRL3_01334 cluster	NRRL3_01334	An13g03040	non-ribosomal peptide synthetase	C-Rpt-A-T-C
NRRL3_01334 cluster	NRRL3_01335	An13g03030	fungal-specific transcription factor	
NRRL3_01339 cluster	NRRL3_01336	An13g03000	cytochrome P450	
NRRL3_01339 cluster	NRRL3_01337	An13g02990	hypothetical protein	
NRRL3_01339 cluster	NRRL3_01338	An13g02980	calycin domain-containing protein	
NRRL3_01339 cluster	NRRL3_01339	An13g02960	hybrid polyketide synthase/non-ribosomal peptide synthetase	KS-AT-(DH-MT-KR-ACP--C-Rpt-A-T-Te)
NRRL3_01339 cluster	NRRL3_01340	An13g02940	enoyl reductase	
NRRL3_01367/NRRL3_01369 cluster	NRRL3_01366	An13g02480	flavin amine oxidase family protein	
NRRL3_01367/NRRL3_01369 cluster	NRRL3_01367	An13g02460	non-ribosomal peptide synthetase-like protein	A-T-Te

NRRL3_01367/NRRL3_01369 cluster	NRRL3_01368	An13g02450	six-hairpin glycosidase-like protein	
NRRL3_01367/NRRL3_01369 cluster	NRRL3_01369	An13g02430	polyketide synthase	KS-AT-DH-ER-KR-ACP
NRRL3_01367/NRRL3_01369 cluster	NRRL3_01370	An13g02420	serine hydrolase FSH family protein	
NRRL3_01367/NRRL3_01369 cluster	NRRL3_01371	An13g02410	magnesium transporter CorA/zinc transporter ZntB family protein	
NRRL3_01367/NRRL3_01369 cluster	NRRL3_01372	An13g02400	fungal-specific transcription factor	
NRRL3_01367/NRRL3_01369 cluster	NRRL3_01373	An13g02390	purine permease	
NRRL3_01367/NRRL3_01369 cluster	NRRL3_01374	NFG	polyamine transporter	
NRRL3_01418 cluster	NRRL3_01414	An13g01890	S-adenosyl-L-methionine dependent methyltransferase domain-containing protein	
NRRL3_01418 cluster	NRRL3_01415	An13g01880 / An08g11860	esterase EstA	
NRRL3_01418 cluster	NRRL3_01416	An13g01870	hypothetical protein	
NRRL3_01418 cluster	NRRL3_01417	An13g01860	amino acid/polyamine transporter I family protein	
NRRL3_01418 cluster	NRRL3_01418	An13g01840	4-dimethylallyltryptophan synthase	PT-DMATS_CynD
NRRL3_01804 cluster	NRRL3_01804	An01g02030	polyketide synthase	(KS-AT-DH)-MT-ER-KR-ACP
NRRL3_01804 cluster	NRRL3_01805	NFG	fatty acyl-CoA reductase-like protein	
NRRL3_01804 cluster	NRRL3_01806	partially NFG	hypothetical protein	
NRRL3_01804 cluster	NRRL3_01807	An01g02040	chloramphenicol acetyltransferase-like protein	
NRRL3_01804 cluster	NRRL3_01808	An01g02050	fungal-specific transcription factor	
NRRL3_01804 cluster	NRRL3_01809	An01g02060	major facilitator superfamily protein	
NRRL3_02191 cluster	NRRL3_02191	An01g06950	polyketide synthase	KS-AT-DH-KR-ER-KR-ACP
NRRL3_02191 cluster	NRRL3_02192	An01g06960	HAD hydrolase subfamily IA protein	
NRRL3_02593/NRRL3_02596 cluster	NRRL3_02592	An01g11760	major facilitator superfamily protein	
NRRL3_02593/NRRL3_02596 cluster	NRRL3_02593	An01g11770	non-ribosomal peptide synthetase-like protein	A-T
NRRL3_02593/NRRL3_02596 cluster	NRRL3_02594	NFG	tyrosinase copper-binding domain-containing protein	
NRRL3_02593/NRRL3_02596 cluster	NRRL3_02595	An01g11780	NUDIX hydrolase domain-containing protein	
NRRL3_02593/NRRL3_02596 cluster	NRRL3_02596	An01g11790	non-ribosomal peptide synthetase-like protein	A-T-Te
NRRL3_02593/NRRL3_02596 cluster	NRRL3_02597	An01g11800	FAD-binding domain-containing protein	
NRRL3_02620/NRRL3_02621 cluster	NRRL3_02619	An01g12020	glycosyltransferase family 2 protein	
NRRL3_02620/NRRL3_02621 cluster	NRRL3_02620	An01g12030	acyl carrier protein	ACP
NRRL3_02620/NRRL3_02621 cluster	NRRL3_02621	An01g12040	polyketide synthase-like protein	KR-ACP
NRRL3_02852 cluster	NRRL3_02844	An01g14760	FAD-binding domain-containing protein	
NRRL3_02852 cluster	NRRL3_02845	An01g14770	fungal-specific transcription factor	

NRRL3_02852 cluster	NRRL3_02846	An01g14780	zinc-type alcohol dehydrogenase superfamily protein	
NRRL3_02852 cluster	NRRL3_02847	An01g14790	maltose/galactoside acetyltransferase domain-containing protein	
NRRL3_02852 cluster	NRRL3_02848	An01g14800	tannase/feruloyl esterase family protein	
NRRL3_02852 cluster	NRRL3_02849	An01g14810	hypothetical protein	
NRRL3_02852 cluster	NRRL3_02850	An01g14820	hypothetical protein	
NRRL3_02852 cluster	NRRL3_02851	An01g14840	alpha/beta hydrolase fold-1 domain-containing protein	
NRRL3_02852 cluster	NRRL3_02852	An01g14850	non-ribosomal peptide synthetase-like protein	A-T-(NADB-AAA)
NRRL3_02852 cluster	NRRL3_02853	An01g14870	hypothetical protein	
NRRL3_02852 cluster	NRRL3_02854	An01g14880	alcohol dehydrogenase [NADP(+)]	
NRRL3_02852 cluster	NRRL3_02855	An01g14890	kynurenine formamidase-like protein	
NRRL3_02852 cluster	NRRL3_02856	An01g14900	major facilitator superfamily protein	
NRRL3_02852 cluster	NRRL3_02857	An01g14910	fungal-specific transcription factor	
NRRL3_02852 cluster	NRRL3_02858	An01g14920	peptidase M24 domain-containing protein	
NRRL3_02852 cluster	NRRL3_02859	An01g14930	major facilitator sugar transporter-like family protein	
NRRL3_02887 cluster	NRRL3_02887	An12g10860	non-ribosomal peptide synthetase-like protein	A-T-Te
NRRL3_02887 cluster	NRRL3_02888	An12g10850	HpcH/HpaI aldolase/citrate lyase domain-containing protein	
NRRL3_02887 cluster	NRRL3_02889	An12g10840	hypothetical protein	
NRRL3_02887 cluster	NRRL3_02890	An12g10830	RmlC-like cupin domain-containing protein	
NRRL3_02887 cluster	NRRL3_02891	An12g10810	hypothetical protein	
NRRL3_02887 cluster	NRRL3_02892	An12g10790	chloroperoxidase domain-containing protein	
NRRL3_02961.1 cluster	NRRL3_02955	An12g10150	YjgF/YER057c/UK114 family protein	
NRRL3_02961.1 cluster	NRRL3_02956	An12g10140	hydroxyacid oxidase	
NRRL3_02961.1 cluster	NRRL3_02957	An12g10130	amino acid/polyamine transporter I family protein	
NRRL3_02961.1 cluster	NRRL3_02958	An12g10120	class II aldolase/adducin family protein	
NRRL3_02961.1 cluster	NRRL3_02959	An12g10110	fungal-specific transcription factor	
NRRL3_02961.1 cluster	NRRL3_02960	An12g10100	ketopantoate reductase ApbA/PANE-like protein	
NRRL3_02961.1 cluster	NRRL3_02961.1	An12g10090	non-ribosomal peptide synthetase-like protein	A-T-Te
NRRL3_02961.1 cluster	NRRL3_02962	An12g10080	uncharacterized protein	
NRRL3_02961.1 cluster	NRRL3_02963	An12g10060	short-chain dehydrogenase/reductase family protein	
NRRL3_02961.1 cluster	NRRL3_02964	An12g10050	amidohydrolase-related domain-containing protein	
NRRL3_03167 cluster	NRRL3_03164	An12g07260	alpha/beta hydrolase fold-1 domain-containing protein	
NRRL3_03167 cluster	NRRL3_03165	An12g07250	amidohydrolase-related domain-containing protein	

NRRL3_03167 cluster	NRRL3_03166	An12g07240	P-loop containing nucleoside triphosphate hydrolase domain-containing protein	
NRRL3_03167 cluster	NRRL3_03167	An12g07230	non-ribosomal peptide synthetase	A-T-C-Rpt-A-MT-A-T-C-A-T-Te
NRRL3_03167 cluster	NRRL3_03168	not called	hypothetical protein	
NRRL3_03167 cluster	NRRL3_03169	An12g07220	NAD(P)-binding domain-containing protein	
NRRL3_03167 cluster	NRRL3_03170	An12g07200	hypothetical protein	
NRRL3_03167 cluster	NRRL3_03171	An12g07290	fungal-specific transcription factor	
NRRL3_03167 cluster	NRRL3_03172	An12g07280	alpha/beta hydrolase fold-1 domain-containing protein	
NRRL3_03184 cluster	NRRL3_03180	An12g07110	salicylate synthase-like protein	
NRRL3_03184 cluster	NRRL3_03181	not called	alkyl-hydroperoxide reductase D-like protein	
NRRL3_03184 cluster	NRRL3_03182	An12g07100	hypothetical protein	
NRRL3_03184 cluster	NRRL3_03183	An12g07090	FAD-binding domain-containing protein	
NRRL3_03184 cluster	NRRL3_03184	An12g07070	polyketide synthase	KS-AT-DH-ER-KR-ACP
NRRL3_03184 cluster	NRRL3_03185	An12g07060	serine hydrolase FSH family protein	
NRRL3_03756 cluster	NRRL3_03750	An15g02070	FAD-binding domain-containing protein	
NRRL3_03756 cluster	NRRL3_03751	An15g02080	fungal-specific transcription factor	
NRRL3_03756 cluster	NRRL3_03752	not called	carbohydrate-binding module family 16 protein	
NRRL3_03756 cluster	NRRL3_03753	An15g02090	hypothetical protein	
NRRL3_03756 cluster	NRRL3_03754	An15g02110	tannase/feruloyl esterase family protein	
NRRL3_03756 cluster	NRRL3_03755	An15g02120	two-component system protein	
NRRL3_03756 cluster	NRRL3_03756	An15g02130	polyketide synthase	KS-AT-DH-MT-ER-KR-ACP
NRRL3_03756 cluster	NRRL3_03757	An15g02140	serine hydrolase FSH family protein	
NRRL3_03756 cluster	NRRL3_03758	An15g02150	hypothetical protein	
NRRL3_03756 cluster	NRRL3_03759	An15g02160	carboxymuconolactone decarboxylase-like protein	
NRRL3_03756 cluster	NRRL3_03760	An15g02180	FAD-binding domain-containing protein	
NRRL3_03756 cluster	NRRL3_03761	An15g02190	major facilitator superfamily protein	
NRRL3_03756 cluster	NRRL3_03762	An15g02200	GMC oxidoreductase	
NRRL3_03756 cluster	NRRL3_03762	An15g02200	GMC oxidoreductase	
NRRL3_03891 cluster	NRRL3_03891	An15g04140	polyketide synthase	KS-AT-DH-MT-ER-KR-ACP
NRRL3_03891 cluster	NRRL3_03892	An15g04150	short-chain dehydrogenase/reductase family protein	
NRRL3_03891 cluster	NRRL3_03893	An15g04160 An15g04170	six-bladed beta-propeller, TolB-like protein	

NRRL3_03891 cluster	NRRL3_03894	An15g04180	cytochrome P450 family protein	
NRRL3_03891 cluster	NRRL3_03895	An15g04190	hypothetical protein	
NRRL3_03891 cluster	NRRL3_03896	An15g04200	hypothetical protein	
NRRL3_03891 cluster	NRRL3_03897	An15g04210	aldo/keto reductase/potassium channel subunit beta family protein	
NRRL3_03891 cluster	NRRL3_03898	An15g04220	fungal-specific transcription factor	
NRRL3_03977 cluster	NRRL3_03972	An15g05040	beta-lactamase-related protein	
NRRL3_03977 cluster	NRRL3_03973	An15g05050	nucleoside phosphorylase family protein	
NRRL3_03977 cluster	NRRL3_03974	An15g05060	major facilitator superfamily protein	
NRRL3_03977 cluster	NRRL3_03975	An15g05070	cytochrome P450 family protein	
NRRL3_03977 cluster	NRRL3_03976	An15g05080	fungal-specific transcription factor	
NRRL3_03977 cluster	NRRL3_03977	An15g05090	polyketide synthase	KS-AT-DH-MT-ER-KR-ACP
NRRL3_03977 cluster	NRRL3_03978	An15g05100	fungal-specific transcription factor	
NRRL3_03977 cluster	NRRL3_03979	An15g05110	cytochrome P450	
NRRL3_04180 cluster	NRRL3_04173	An15g07460	oligopeptide transporter OPT superfamily protein	
NRRL3_04180 cluster	NRRL3_04174	An15g07470	hypothetical protein	
NRRL3_04180 cluster	NRRL3_04175	An15g07480	hypothetical protein	
NRRL3_04180 cluster	NRRL3_04176	An15g07490	TauD/TfdA-like domain-containing protein	
NRRL3_04180 cluster	NRRL3_04177	An15g07500	ribokinase/fructokinase family protein	
NRRL3_04180 cluster	NRRL3_04178	An15g07510	proton-dependent oligopeptide transporter family protein	
NRRL3_04180 cluster	NRRL3_04179	An15g07520	uncharacterized protein	
NRRL3_04180 cluster	NRRL3_04180	An15g07530	non-ribosomal peptide synthetase	A-T-C-A-T-C-A-T-C-A-T-C-A-T-C
NRRL3_04226 cluster	NRRL3_04223	assembly issue	alpha-hydroxy acid dehydrogenase family protein, FMN-dependent	
NRRL3_04226 cluster	NRRL3_04224	assembly issue	hypothetical protein	
NRRL3_04226 cluster	NRRL3_04225	assembly issue	flavin monooxygenase-like protein	
NRRL3_04226 cluster	NRRL3_04226	assembly issue	hybrid polyketide synthase/non-ribosomal peptide synthetase	KS-AT-DH-KR-ACP--C-A-T-Te
NRRL3_04226 cluster	NRRL3_04227	An07g00010	EthD domain-containing protein	
NRRL3_04226 cluster	NRRL3_04228	An07g00020	alpha/beta hydrolase fold-5 domain-containing protein	
NRRL3_04226 cluster	NRRL3_04229	An07g00030	aldo/keto reductase/potassium channel subunit beta family protein	
NRRL3_04226 cluster	NRRL3_04230	An07g00040	AMP-dependent synthetase/ligase domain-containing protein	
NRRL3_04226 cluster	NRRL3_04231	An07g00050	fungal-specific transcription factor	

NRRL3_04226 cluster	NRRL3_04232	An07g00060	major facilitator superfamily protein	
NRRL3_04226 cluster	NRRL3_04233	An07g00070	EthD domain-containing protein	
NRRL3_04226 cluster	NRRL3_04234	An07g00080	cytochrome P450	
NRRL3_04226 cluster	NRRL3_04235	An07g00090	amidohydrolase-related domain-containing protein	
NRRL3_04226 cluster	NRRL3_04236	An07g00100	amidase family protein	
NRRL3_04226 cluster	NRRL3_04237	An07g00110	beta-lactamase-related protein	
NRRL3_04226 cluster	NRRL3_04238	An07g00130	GNAT domain-containing protein	
NRRL3_04226 cluster	NRRL3_04239	An07g00150	major facilitator superfamily protein	
NRRL3_04226 cluster	NRRL3_04240	An07g00170	hypothetical protein	
NRRL3_04226 cluster	NRRL3_04241	An07g00190	S-adenosyl-L-methionine dependent methyltransferase domain-containing protein	
NRRL3_04226 cluster	NRRL3_04242	An07g00200	alpha/beta hydrolase fold-1 domain-containing protein	
NRRL3_04305 cluster	NRRL3_04305	An07g01030	polyketide synthase	KS-AT-DH-(ACP)
NRRL3_04420 cluster	NRRL3_04419	An07g02550	cytochrome P450 family protein	
NRRL3_04420 cluster	NRRL3_04420	An07g02560	4-dimethylallyltryptophan synthase	PT-DMATS_CynD
NRRL3_04420 cluster	NRRL3_04421	An07g02540	MFS-type short-chain carboxylic acid transporter	
NRRL3_05440 cluster	NRRL3_05437	An02g10170	major facilitator superfamily protein	
NRRL3_05440 cluster	NRRL3_05438	An02g10160	aminotransferase class I and II family protein	
NRRL3_05440 cluster	NRRL3_05439	An02g10150	CoA-transferase family III protein	
NRRL3_05440 cluster	NRRL3_05440	An02g10140	non-ribosomal peptide synthetase-like protein	A-T-Te
NRRL3_05440 cluster	NRRL3_05441	NFG	fungal-specific transcription factor	
NRRL3_05484 cluster	NRRL3_05484	An02g09430	polyketide synthase	KS-AT-DH-MT-ER-KR-ACP
NRRL3_05484 cluster	NRRL3_05485	An02g09420	cytochrome P450 family protein	
NRRL3_05588 cluster	NRRL3_05580	An02g08370	flavin amine oxidase family protein	
NRRL3_05588 cluster	NRRL3_05581	An02g08360	RmlC-like cupin domain-containing protein	
NRRL3_05588 cluster	NRRL3_05582	An02g08350	fungal-specific transcription factor	
NRRL3_05588 cluster	NRRL3_05583	An02g08340	amino acid/polyamine transporter I family protein	
NRRL3_05588 cluster	NRRL3_05584	An02g08330	major facilitator superfamily protein	
NRRL3_05588 cluster	NRRL3_05585	An02g08320	hypothetical protein	
NRRL3_05588 cluster	NRRL3_05586	An02g08310	enoyl reductase	
NRRL3_05588 cluster	NRRL3_05587	An02g08300	EthD domain-containing protein	
NRRL3_05588 cluster	NRRL3_05588	An02g08290	hybrid polyketide synthase/non-ribosomal peptide synthetase	KS-AT-DH-MT-KR-ACP--C-A-T-E

NRRL3_05848 cluster	NRRL3_05846	An02g05090	cytochrome P450 family protein	
NRRL3_05848 cluster	NRRL3_05847	An02g05080	FAD-binding domain-containing protein	
NRRL3_05848 cluster	NRRL3_05848	An02g05070	non-ribosomal peptide synthetase	A-T-C-T-C
NRRL3_06189 cluster	NRRL3_06189	An02g00840	non-ribosomal peptide synthetase-like protein	A-T-Te
NRRL3_06189 cluster	NRRL3_06190	An02g00810	COMT-type O-methyltransferase family protein	
NRRL3_06189 cluster	NRRL3_06191	An02g00800	hypothetical protein	
NRRL3_06189 cluster	NRRL3_06192	An02g00790	hypothetical protein	
NRRL3_06189 cluster	NRRL3_06193	An02g00780	alpha/beta hydrolase fold domain-containing protein	
NRRL3_06189 cluster	NRRL3_06194	An02g00760	major facilitator superfamily protein	
NRRL3_06189 cluster	NRRL3_06195	An02g00750	hypothetical protein	
NRRL3_06189 cluster	NRRL3_06196	An02g00740	FAD-binding domain-containing protein	
NRRL3_06217 cluster	NRRL3_06216	An02g00460	FAD-binding domain-containing protein	
NRRL3_06217 cluster	NRRL3_06217	An02g00450	polyketide synthase	KS-AT-DH-MT-ER-KR-ACP
NRRL3_06217 cluster	NRRL3_06218	An02g00440	serine hydrolase FSH family protein	
NRRL3_06217 cluster	NRRL3_06219	An02g00420	major facilitator superfamily protein	
NRRL3_06237 cluster	NRRL3_06224	An02g00340	amidase family protein	
NRRL3_06237 cluster	NRRL3_06225	An02g00330	hypothetical protein	
NRRL3_06237 cluster	NRRL3_06226	An02g00320	hypothetical protein	
NRRL3_06237 cluster	NRRL3_06227	An02g00310	aldo-keto reductase family protein	
NRRL3_06237 cluster	NRRL3_06228	An02g00300	dienelactone hydrolase domain-containing protein	
NRRL3_06237 cluster	NRRL3_06229	An02g00290	fungal-specific transcription factor	
NRRL3_06237 cluster	NRRL3_06230	An02g00280	major facilitator superfamily domain-containing protein	
NRRL3_06237 cluster	NRRL3_06231	An02g00270	RmlC-like cupin domain-containing protein	
NRRL3_06237 cluster	NRRL3_06232	An02g00260	fungal-specific transcription factor	
NRRL3_06237 cluster	NRRL3_06233	An02g00250	deoxyribose-phosphate aldolase	
NRRL3_06237 cluster	NRRL3_06234	An02g00240	amino acid/polyamine transporter I family protein	
NRRL3_06237 cluster	NRRL3_06235	An02g00230	phytanoyl-CoA dioxygenase family protein	
NRRL3_06237 cluster	NRRL3_06236	An02g00220	short-chain dehydrogenase/reductase family protein	
NRRL3_06237 cluster	NRRL3_06237	An02g00210	non-ribosomal peptide synthetase-like protein	A-T-Dhe
NRRL3_06237 cluster	NRRL3_06238	An02g00200	hypothetical protein	
NRRL3_06237 cluster	NRRL3_06239	An02g00190	amidase family protein	
NRRL3_06340/NRRL3_06341 cluster	NRRL3_06340	An10g00630	fatty acid synthase alpha subunit	MPT-ACP-KR-KS
NRRL3_06340/NRRL3_06341 cluster	NRRL3_06341	An10g00650	fatty acid synthase beta subunit	AT-ER-DH-MPT

NRRL3_06340/NRRL3_06341 cluster	NRRL3_06342	An10g00660	cytochrome P450 family protein	
NRRL3_06340/NRRL3_06341 cluster	NRRL3_06343	An10g00670	methyltransferase type 11 domain-containing protein	
NRRL3_06340/NRRL3_06341 cluster	NRRL3_06344	An10g00680	ATPase, V0 complex, proteolipid subunit c	
NRRL3_06340/NRRL3_06341 cluster	NRRL3_06345	An10g00690	major facilitator superfamily protein	
NRRL3_06340/NRRL3_06341 cluster	NRRL3_06346	An10g00700	drug resistance protein	
NRRL3_06432 cluster	NRRL3_06431	An17g00440	cytochrome P450 family protein	
NRRL3_06432 cluster	NRRL3_06432	An17g00450	sesquiterpene cyclase	Isoprenoid_Biosyn_C1
NRRL3_06801 cluster	NRRL3_06796	NFG	aminoglycoside phosphotransferase domain-containing protein	
NRRL3_06801 cluster	NRRL3_06797	NFG	RTA-like protein	
NRRL3_06801 cluster	NRRL3_06798	NFG	RTA-like protein	
NRRL3_06801 cluster	NRRL3_06799	An16g06750	beta-lactamase-related protein	
NRRL3_06801 cluster	NRRL3_06800	An16g06740	ABC transporter type 1, transmembrane domain-containing protein	
NRRL3_06801 cluster	NRRL3_06801	An16g06720	non-ribosomal peptide synthetase	A-C-T-MT-A-T-C-A-T-C-A-T-C-A-T-C
NRRL3_07380 cluster	NRRL3_07368	An16g00750	fungal-specific transcription factor	
NRRL3_07380 cluster	NRRL3_07369	An16g00740	NmrA-like family protein	
NRRL3_07380 cluster	NRRL3_07370	An16g00730	major facilitator superfamily protein	
NRRL3_07380 cluster	NRRL3_07371	An16g00720	major facilitator superfamily protein	
NRRL3_07380 cluster	NRRL3_07372	An16g00710	caleosin-related protein	
NRRL3_07380 cluster	NRRL3_07373	An16g00700	FAD-dependent oxidoreductase domain-containing protein	
NRRL3_07380 cluster	NRRL3_07374	An16g00680	RTA-like protein	
NRRL3_07380 cluster	NRRL3_07375	An16g00670	hypothetical protein	
NRRL3_07380 cluster	NRRL3_07376	An16g00660	NAD(P)-binding domain-containing protein	
NRRL3_07380 cluster	NRRL3_07377	An16g00650	major facilitator superfamily protein	
NRRL3_07380 cluster	NRRL3_07378	An16g00630	chloroperoxidase domain-containing protein	
NRRL3_07380 cluster	NRRL3_07379	An16g00620	hypothetical protein	
NRRL3_07380 cluster	NRRL3_07380	An16g00600	non-ribosomal peptide synthetase-like protein	A-T-E
NRRL3_07380 cluster	NRRL3_07381	An16g00550	arylacetonitrilase NitAn	
NRRL3_07443 cluster	NRRL3_07443	An04g10030	polyketide synthase	KS-AT-DH-MT-ER-KR-ACP
NRRL3_07443 cluster	NRRL3_07444	An04g10020	serine hydrolase FSH family protein	
NRRL3_07443 cluster	NRRL3_07445	An04g10010	hypothetical protein	
NRRL3_07443 cluster	NRRL3_07446	An04g10000	major facilitator superfamily protein	

NRRL3_07443 cluster	NRRL3_07447	An04g09990	short-chain dehydrogenase/reductase family protein	
NRRL3_07739 cluster	NRRL3_07736	An04g06290	fungal-specific transcription factor	
NRRL3_07739 cluster	NRRL3_07737	An04g06280	alkyl-hydroperoxide reductase D-like protein	
NRRL3_07739 cluster	NRRL3_07738	An04g06270	methyltransferase domain-containing protein	
NRRL3_07739 cluster	NRRL3_07739	An04g06260	non-ribosomal peptide synthetase	A-T-C-A-T-C
NRRL3_07739 cluster	NRRL3_07740	An04g06250	major facilitator superfamily protein	
NRRL3_07739 cluster	NRRL3_07741	An04g06240	fungal-specific transcription factor	
NRRL3_07812 cluster	NRRL3_07810	An04g05440	xanthine dehydrogenase	
NRRL3_07812 cluster	NRRL3_07811	An04g05430	hypothetical protein	
NRRL3_07812 cluster	NRRL3_07812	An04g05420	non-ribosomal peptide synthetase-like protein	A-T-NADB
NRRL3_07812 cluster	NRRL3_07813	An04g05410	hypothetical protein	
NRRL3_07812 cluster	NRRL3_07814	An04g05380	versiconal hemiacetal acetate reductase	
NRRL3_07881/NRRL3_07884 cluster	NRRL3_07873	An04g04490	fungal-specific transcription factor	
NRRL3_07881/NRRL3_07884 cluster	NRRL3_07874	An04g04480	zinc-type alcohol dehydrogenase superfamily protein	
NRRL3_07881/NRRL3_07884 cluster	NRRL3_07875	An04g04470	GNAT domain-containing protein	
NRRL3_07881/NRRL3_07884 cluster	NRRL3_07876	An04g04460	hypothetical protein	
NRRL3_07881/NRRL3_07884 cluster	NRRL3_07877	An04g04440	FAD-binding domain-containing protein	
NRRL3_07881/NRRL3_07884 cluster	NRRL3_07878	An04g04430	tannase/feruloyl esterase family protein	
NRRL3_07881/NRRL3_07884 cluster	NRRL3_07879	An04g04400	ankyrin repeat domain-containing protein	
NRRL3_07881/NRRL3_07884 cluster	NRRL3_07880	An04g04390	isochorismatase family protein	
NRRL3_07881/NRRL3_07884 cluster	NRRL3_07881	An04g04380	non-ribosomal peptide synthetase-like protein	A-T-Te
NRRL3_07881/NRRL3_07884 cluster	NRRL3_07882	An04g04370	phenylalanine ammonia-lyase	
NRRL3_07881/NRRL3_07884 cluster	NRRL3_07883	An04g04360	thioesterase domain-containing protein	
NRRL3_07881/NRRL3_07884 cluster	NRRL3_07884	An04g04340	polyketide synthase	KS-AT-DH-MT-KR-ACP
NRRL3_07881/NRRL3_07884 cluster	NRRL3_07885	An04g04330	AMP-dependent synthetase/ligase domain-containing protein	
NRRL3_08167 cluster	NRRL3_08166	An04g01160	UbiB domain-containing protein	
NRRL3_08167 cluster	NRRL3_08167	An04g01150	non-ribosomal peptide synthetase-like protein	A-T-Te
NRRL3_08167 cluster	NRRL3_08168	An04g01140	aminotransferase, class I and II family protein	
NRRL3_08167 cluster	NRRL3_08169	An04g01130	cytokinesis protein SepH	
NRRL3_08167 cluster	NRRL3_08170	An04g01120	folylpolyglutamate synthetase-like protein	
NRRL3_08318 cluster	NRRL3_08313	An03g06430	NADP-dependent alcohol dehydrogenase	
NRRL3_08318 cluster	NRRL3_08314	An03g06420	GNAT domain-containing protein	
NRRL3_08318 cluster	NRRL3_08315	An03g06410	methylsterol monooxygenase	

NRRL3_08318 cluster	NRRL3_08316	An03g06400	fungal-specific transcription factor	
NRRL3_08318 cluster	NRRL3_08317	An03g06390	FAD-binding domain-containing protein	
NRRL3_08318 cluster	NRRL3_08318	An03g06380	polyketide synthase	KS-AT-DH-MT-ER-KR-ACP
NRRL3_08318 cluster	NRRL3_08319	An03g06370	fungal-specific transcription factor	
NRRL3_08341 cluster	NRRL3_08340	An03g06020	carboxymuconolactone decarboxylase-like protein	
NRRL3_08341 cluster	NRRL3_08341	An03g06010	non-ribosomal peptide synthetase	A-T-C-A-T-NADB
NRRL3_08341 cluster	NRRL3_08342	An03g06000	FAD-binding domain-containing protein	
NRRL3_08341 cluster	NRRL3_08343	An03g05990	COMT-type O-methyltransferase family protein	
NRRL3_08341 cluster	NRRL3_08344	An03g05980	fungal-specific transcription factor	
NRRL3_08341 cluster	NRRL3_08345	An03g05960	epoxide hydrolase-like protein	
NRRL3_08341 cluster	NRRL3_08346	An03g05950	major facilitator superfamily protein	
NRRL3_08369 cluster	NRRL3_08368	An03g05690	ankyrin repeat domain-containing protein	
NRRL3_08369 cluster	NRRL3_08369	An03g05680	non-ribosomal peptide synthetase-like protein	A-T-Te
NRRL3_08369 cluster	NRRL3_08370	An03g05670	GMC oxidoreductase	
NRRL3_08369 cluster	NRRL3_08371	An03g05660	catalase	
NRRL3_08369 cluster	NRRL3_08372	An03g05650	fungal-specific transcription factor	
NRRL3_08388 cluster	NRRL3_08381	An03g05500	fungal-specific transcription factor	
NRRL3_08388 cluster	NRRL3_08382	An03g05490	iron-type alcohol dehydrogenase domain-containing protein	
NRRL3_08388 cluster	NRRL3_08383	An03g05480	COMT-type O-methyltransferase family protein	
NRRL3_08388 cluster	NRRL3_08384	An03g05470	RmlC-like cupin domain-containing protein	
NRRL3_08388 cluster	NRRL3_08385	An03g05460	cytochrome P450 family protein	
NRRL3_08388 cluster	NRRL3_08386	An03g05450	FAD-binding domain-containing protein	
NRRL3_08388 cluster	NRRL3_08387	An03g05450	EthD domain-containing protein	
NRRL3_08388 cluster	NRRL3_08388	An03g05440	polyketide synthase	SAT-KS-AT-PT-ACP-ACP-Te
NRRL3_08388 cluster	NRRL3_08389	An03g05430	COMT-type O-methyltransferase family protein	
NRRL3_08411 cluster	NRRL3_08411	An03g05140	polyketide synthase	KS-AT-DH-MT-ER-KR-ACP
NRRL3_08411 cluster	NRRL3_08412	An03g05130	serine hydrolase FSH family protein	
NRRL3_08411 cluster	NRRL3_08413	An03g05120	major facilitator superfamily protein	
NRRL3_08411 cluster	NRRL3_08414	An03g05110	serine hydrolase FSH family protein	
NRRL3_08647 cluster	NRRL3_08643	An03g01860	fungal-specific transcription factor	
NRRL3_08647 cluster	NRRL3_08644	An03g01850	isochorismatase family protein	
NRRL3_08647 cluster	NRRL3_08645	An03g01840	non-haem dioxygenase N-terminal domain-containing protein	
NRRL3_08647 cluster	NRRL3_08646	An03g01830	hypothetical protein	

NRRL3_08647 cluster	NRRL3_08647	An03g01820	polyketide synthase	SAT-KS-AT-ACP-Te
NRRL3_08647 cluster	NRRL3_08648	An03g01810	serine hydrolase FSH family protein	
NRRL3_08647 cluster	NRRL3_08649	An03g01800	acyl-CoA-binding protein	
NRRL3_08647 cluster	NRRL3_08650	An03g01790	major facilitator superfamily protein	
NRRL3_08729 cluster	NRRL3_08721	An03g00750	fungal-specific transcription factor	
NRRL3_08729 cluster	NRRL3_08722	An03g00740	mannan endo-1,6-alpha-mannosidase	
NRRL3_08729 cluster	NRRL3_08723	An03g00730	copper amine oxidase	
NRRL3_08729 cluster	NRRL3_08724	An03g00720	hypothetical protein	
NRRL3_08729 cluster	NRRL3_08725	An03g00690	hypothetical protein	
NRRL3_08729 cluster	NRRL3_08726	An03g00680	major facilitator superfamily protein	
NRRL3_08729 cluster	NRRL3_08727	An03g00670	opine dehydrogenase domain-containing protein	
NRRL3_08729 cluster	NRRL3_08728	An03g00660	taurine dioxygenase	
NRRL3_08729 cluster	NRRL3_08729	An03g00650	non-ribosomal peptide synthetase	A-T-C-A-T-C
NRRL3_08729 cluster	NRRL3_08730	An03g00640	amino acid transporter, transmembrane domain-containing protein BrsA-25	
NRRL3_08732 cluster	NRRL3_08732	An03g00590	sesquiterpene cyclase	Isoprenoid_Biosyn_C1
NRRL3_08732 cluster	NRRL3_08733	An03g00580	cytochrome P450 family protein	
NRRL3_08732 cluster	NRRL3_08734	An03g00560	hypothetical protein	
NRRL3_08732 cluster	NRRL3_08735	An03g00550	hypothetical protein	
NRRL3_08732 cluster	NRRL3_08736	An03g00530	FAD-binding domain-containing protein	
NRRL3_08775 cluster	NRRL3_08768	An03g00180	cytochrome P450 family protein	
NRRL3_08775 cluster	NRRL3_08769	An03g00170	major facilitator superfamily protein	
NRRL3_08775 cluster	NRRL3_08770	An03g00160	fungal-specific transcription factor	
NRRL3_08775 cluster	NRRL3_08771	An03g00150	hypothetical protein	
NRRL3_08775 cluster	NRRL3_08772	An03g00140	zinc-type alcohol dehydrogenase superfamily protein	
NRRL3_08775 cluster	NRRL3_08773	An03g00130	GMC oxidoreductase	
NRRL3_08775 cluster	NRRL3_08774	NFG	hypothetical protein	
NRRL3_08775 cluster	NRRL3_08775	NFG	polyketide synthase	SAT-KS-AT-PT-ACP-ACP-MT-Te
NRRL3_08790 cluster	NRRL3_08788	NFG	alpha/beta hydrolase fold-1 domain-containing protein	
NRRL3_08790 cluster	NRRL3_08789	NFG	major facilitator superfamily protein	
NRRL3_08790 cluster	NRRL3_08790	NFG	non-ribosomal peptide synthetase	A-T-C-A-T-C-A-T-C-A-T-C-A-T-C

NRRL3_08891 cluster	NRRL3_08890	An05g01050	fungal-specific transcription factor	
NRRL3_08891 cluster	NRRL3_08891	An05g01060	non-ribosomal peptide synthetase	A-T-C-C-A-T-C-A-T-C-A-T-C-C
NRRL3_08891 cluster	NRRL3_08892	An05g01070	RTA-like protein	
NRRL3_08891 cluster	NRRL3_08893	An05g01080	ankyrin repeat domain-containing protein	
NRRL3_08891 cluster	NRRL3_08894	not called	HotDog domain-containing protein	
NRRL3_08891 cluster	NRRL3_08895	An05g01100	branched-chain amino acid aminotransferase-like protein	
NRRL3_08891 cluster	NRRL3_08896	An05g01110	ABC transporter type 1, transmembrane domain-containing protein	
NRRL3_08891 cluster	NRRL3_08897	An05g01120	cytochrome P450 family protein	
NRRL3_08969 cluster	NRRL3_08964	An12g02890	aminotransferase class IV family protein	
NRRL3_08969 cluster	NRRL3_08965.1	An12g02880	fungal-specific transcription factor	
NRRL3_08969 cluster	NRRL3_08966	not called	major facilitator superfamily protein	
NRRL3_08969 cluster	NRRL3_08967	An12g02860	major facilitator superfamily protein	
NRRL3_08969 cluster	NRRL3_08968	An12g02850	pyridine nucleotide-disulphide oxidoreductase family protein	
NRRL3_08969 cluster	NRRL3_08969	An12g02840	non-ribosomal peptide synthetase	A-T-C-A-T-C-A-T-C-C-A-T-C
NRRL3_08969 cluster	NRRL3_08970	An12g02830	methyltransferase domain-containing protein	
NRRL3_08969 cluster	NRRL3_08971	An12g02820	major facilitator superfamily protein	
NRRL3_08969 cluster	NRRL3_08972	An12g02810	COMT-type O-methyltransferase family protein	
NRRL3_08969 cluster	NRRL3_08973	An12g02800	major facilitator superfamily protein	
NRRL3_08978/NRRL3_08980/NRRL3_08984.1 cluster	NRRL3_08978	An12g02750	non-ribosomal peptide synthetase	A-T-C
NRRL3_08978/NRRL3_08980/NRRL3_08984.1 cluster	NRRL3_08979	An12g02740	S-adenosyl-L-methionine dependent methyltransferase domain-containing protein	
NRRL3_08978/NRRL3_08980/NRRL3_08984.1 cluster	NRRL3_08980	An12g02730	polyketide synthase	KS-AT-DH-ER-KR-ACP
NRRL3_08978/NRRL3_08980/NRRL3_08984.1 cluster	NRRL3_08981	An12g02720	serine hydrolase FSH family protein	
NRRL3_08984.1 cluster	NRRL3_08982	An12g02700	2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase	
NRRL3_08984.1 cluster	NRRL3_08983	An12g02680	hypothetical protein	
NRRL3_08984.1 cluster	NRRL3_08984.1	An12g02670	polyketide synthase	KS-AT-DH-MT-ER-KR-ACP-Te
NRRL3_08984.1 cluster	NRRL3_08985	An12g02660	PAN domain-containing protein	

NRRL3_08984.1 cluster	NRRL3_08986	An12g02650	methyltransferase domain-containing protein	
NRRL3_08984.1 cluster	NRRL3_08987	An12g02640	aldo-keto reductase	
NRRL3_08984.1 cluster	NRRL3_08988	An12g02630	NmrA-like family protein	
NRRL3_08984.1 cluster	NRRL3_08989	An12g02620	fungal-specific transcription factor	
NRRL3_09034/NRRL3_09038/NRRL3_09039	NRRL3_09030	An12g02090	fungal-specific transcription factor	
NRRL3_09034/NRRL3_09038/NRRL3_09039	NRRL3_09031	An12g02080	cytochrome P450 family protein	
NRRL3_09034/NRRL3_09038/NRRL3_09039	NRRL3_09032	An12g02070	ribonuclease/ribotoxin domain-containing protein	
NRRL3_09034/NRRL3_09038/NRRL3_09039	NRRL3_09033	An12g02060	uncharacterized protein	
NRRL3_09034/NRRL3_09038/NRRL3_09039	NRRL3_09034	An12g02050	polyketide synthase	SAT-KS-AT-ACP-MT-Te
NRRL3_09034/NRRL3_09038/NRRL3_09039	NRRL3_09035	An12g02040	fungal-specific transcription factor	
NRRL3_09034/NRRL3_09038/NRRL3_09039	NRRL3_09036	An12g02020	transferase family protein	
NRRL3_09034/NRRL3_09038/NRRL3_09039	NRRL3_09037	An12g02000	FAD-binding domain-containing protein	
NRRL3_09034/NRRL3_09038/NRRL3_09039	NRRL3_09038	An12g01990	fatty acid synthase beta subunit	AT-ER-DH-MPT
NRRL3_09034/NRRL3_09038/NRRL3_09039	NRRL3_09039	An12g01980	fatty acid synthase alpha subunit	MPT-ACP-KR-KS-PPT
NRRL3_09034/NRRL3_09038/NRRL3_09039	NRRL3_09040	An12g01970	sensitivity to red light reduced SRR1 domain-containing protein	
NRRL3_09034/NRRL3_09038/NRRL3_09039	NRRL3_09041.1	An12g01960	major facilitator superfamily protein	
NRRL3_09351 cluster	NRRL3_09351	An11g09720	polyketide synthase	KS-AT-DH-MT-KR-ACP
NRRL3_09351 cluster	NRRL3_09352	An11g09710	AMP-dependent synthetase/ligase domain-containing protein	
NRRL3_09351 cluster	NRRL3_09353	An11g09700	uroporphyrin-III C-methyltransferase	
NRRL3_09351 cluster	NRRL3_09354	An11g09690	26S proteasome regulatory complex subunit Rpn5	
NRRL3_09351 cluster	NRRL3_09355	An11g09680	serine hydrolase FSH family protein	
NRRL3_09616 cluster	NRRL3_09616	An11g06460	hybrid polyketide synthase/non-ribosomal peptide synthetase	KS-AT-DH-KR-ACP--C-A-T-Te
NRRL3_09616 cluster	NRRL3_09617	An11g06450	EthD domain-containing protein	
NRRL3_09616 cluster	NRRL3_09618	An11g06440	zinc-type alcohol dehydrogenase superfamily protein	
NRRL3_09616 cluster	NRRL3_09619	An11g06430	hypothetical protein	
NRRL3_09616 cluster	NRRL3_09620	An11g06420	major facilitator superfamily protein	
NRRL3_09686/NRRL3_09693 cluster	NRRL3_09686	An11g05570	polyketide synthase	KS-AT-DH-MT-ER-KR-ACP
NRRL3_09686/NRRL3_09693 cluster	NRRL3_09687	An11g05560	aminotransferase, class I and II family protein	
NRRL3_09686/NRRL3_09693 cluster	NRRL3_09688	An11g05550	major facilitator superfamily protein	
NRRL3_09686/NRRL3_09693 cluster	NRRL3_09689	not called	fungal-specific transcription factor	
NRRL3_09686/NRRL3_09693 cluster	NRRL3_09693	An11g05500	non-ribosomal peptide synthetase-like protein	A-T-Te
NRRL3_09789/NRRL3_09792 cluster	NRRL3_09788	An11g04290	short-chain dehydrogenase/reductase family protein	

NRRL3_09789/NRRL3_09792 cluster	NRRL3_09789	An11g04280	polyketide synthase	KS-AT-DH-ER-KR-ACP
NRRL3_09789/NRRL3_09792 cluster	NRRL3_09790	An11g04270	zinc-type alcohol dehydrogenase superfamily protein	
NRRL3_09789/NRRL3_09792 cluster	NRRL3_09791	An11g04260	serine hydrolase FSH family protein	
NRRL3_09789/NRRL3_09792 cluster	NRRL3_09792	An11g04250	non-ribosomal peptide synthetase-like protein	A-T-Te
NRRL3_09789/NRRL3_09792 cluster	NRRL3_09793	An11g04240	hypothetical protein	
NRRL3_09789/NRRL3_09792 cluster	NRRL3_09794	An11g04220	cytochrome P450	
NRRL3_09827 cluster	NRRL3_09824	An11g03950	alpha/beta hydrolase fold-1 domain-containing protein	
NRRL3_09827 cluster	NRRL3_09825	An11g03940	cytochrome P450 family protein	
NRRL3_09827 cluster	NRRL3_09826	An11g03930	serine hydrolase FSH family protein	
NRRL3_09827 cluster	NRRL3_09827	An11g03920	polyketide synthase	KS-AT-DH-MT-ER-KR-ACP
NRRL3_09848 cluster	NRRL3_09846	An11g03690	fungal-specific transcription factor	
NRRL3_09848 cluster	NRRL3_09847	An11g03680	zinc-type alcohol dehydrogenase superfamily protein	
NRRL3_09848 cluster	NRRL3_09848	An11g03670	polyketide synthase	(KS-AT-DH-MT-ER-KR-ACP)-Te
NRRL3_10148 cluster	NRRL3_10146	An11g00070	COMT-type O-methyltransferase family protein	
NRRL3_10148 cluster	NRRL3_10147	An11g00060	hypothetical protein	
NRRL3_10148 cluster	NRRL3_10148	An11g00050	non-ribosomal peptide synthetase	A-T-T-C-Rpt-C-A-T-C-Rpt-(A)[MT}-A-T-C
NRRL3_10209 cluster	NRRL3_10204	An18g00470	dienelactone hydrolase domain-containing protein	
NRRL3_10209 cluster	NRRL3_10205	An18g00480	major facilitator superfamily protein	
NRRL3_10209 cluster	NRRL3_10206	An18g00490	FAD-binding domain-containing protein	
NRRL3_10209 cluster	NRRL3_10207	An18g00500	cytochrome P450 family protein	
NRRL3_10209 cluster	NRRL3_10208	An18g00510	FAD-binding domain-containing protein	
NRRL3_10209 cluster	NRRL3_10209	An18g00520	hybrid polyketide synthase/non-ribosomal peptide synthetase	KS-AT-DH-KR-ER-KR-ACP--C-Rpt-A-T-Te
NRRL3_10375 cluster	NRRL3_10368	An18g02650	major facilitator superfamily protein	
NRRL3_10375 cluster	NRRL3_10369.1	NFG	methyltransferase type 11 domain-containing protein	
NRRL3_10375 cluster	NRRL3_10370	NFG	fungal-specific transcription factor	
NRRL3_10375 cluster	NRRL3_10371	NFG	serine hydrolase FSH family protein	
NRRL3_10375 cluster	NRRL3_10372	NFG	aldose 1-/glucose-6-phosphate 1-epimerase family protein	
NRRL3_10375 cluster	NRRL3_10373	NFG	cytochrome P450 family protein	
NRRL3_10375 cluster	NRRL3_10374	NFG	enoyl reductase	
NRRL3_10375 cluster	NRRL3_10375	NFG	polyketide synthase	KS-AT-DH-MT-KR-ACP

NRRL3_10375 cluster	NRRL3_10376	An18g02680	cytochrome P450 family protein	
NRRL3_10375 cluster	NRRL3_10377	An18g02690	multicopper oxidase Mco1I	
NRRL3_10375 cluster	NRRL3_10378	An18g02700	cytochrome P450 family protein	
NRRL3_10375 cluster	NRRL3_10379	An18g02710	terpenoid cyclases/protein prenyltransferase alpha-alpha toroid domain-containing protein	
NRRL3_10375 cluster	NRRL3_10380	An18g02720	glutathione S-transferase	
NRRL3_10375 cluster	NRRL3_10381	An18g02730	hypothetical protein	
NRRL3_10375 cluster	NRRL3_10382	An18g02740	cytochrome P450 family protein	
NRRL3_10375 cluster	NRRL3_10383	An18g02750	S-adenosyl-L-methionine dependent methyltransferase domain-containing protein	
NRRL3_10375 cluster	NRRL3_10384	An18g02752	alpha/beta hydrolase fold-5 domain-containing protein	
NRRL3_10912 cluster	NRRL3_10911	An08g02290	polyamine transporter	
NRRL3_10912 cluster	NRRL3_10912	An08g02310	non-ribosomal peptide synthetase	A-T-C-C-A-C-A-T-C-A-T-C-C-T-C-(T)
NRRL3_11121 cluster	NRRL3_11120	An08g04810	enolase superfamily protein	
NRRL3_11121 cluster	NRRL3_11121	An08g04820	non-ribosomal peptide synthetase-like protein	A-T-NADB
NRRL3_11458 cluster	NRRL3_11458	An08g09220	non-ribosomal peptide synthetase-like protein	C-T
NRRL3_11458 cluster	NRRL3_11459	An08g09230	cytochrome P450	
NRRL3_11458 cluster	NRRL3_11460	An08g09240	fungal-specific transcription factor	
NRRL3_11458 cluster	NRRL3_11461	An08g09250	major facilitator superfamily protein	
NRRL3_11458 cluster	NRRL3_11462	An08g09260	acetoacetyl-CoA synthetase	
NRRL3_11458 cluster	NRRL3_11463	An08g09280	fumarylacetoacetate hydrolase family protein	
NRRL3_11726 cluster	NRRL3_11725	An06g00460	carboxylesterase family protein	
NRRL3_11726 cluster	NRRL3_11726	An06g00430	polyketide synthase, type III	KS
NRRL3_11726 cluster	NRRL3_11727	NFG	dihydrolipoamide succinyltransferase	
NRRL3_11726 cluster	NRRL3_11728	An06g00390	2-oxoglutarate dehydrogenase, mitochondrial	
NRRL3_11763/NRRL3_11767 cluster	NRRL3_11759	An08g10970	dienelactone hydrolase domain-containing protein	
NRRL3_11763/NRRL3_11767 cluster	NRRL3_11760	An08g10860	fungal-specific transcription factor	
NRRL3_11763/NRRL3_11767 cluster	NRRL3_11761	An08g10850	major facilitator superfamily protein	
NRRL3_11763/NRRL3_11767 cluster	NRRL3_11762	An08g10840	zinc finger domain-containing protein, RING/FYVE/PHD-type	
NRRL3_11763/NRRL3_11767 cluster	NRRL3_11763	An08g10930	fatty acid synthase alpha subunit	MPT-ACP-KR-KS-PPT
NRRL3_11763/NRRL3_11767 cluster	NRRL3_11764	An08g10920	citrate synthase-like protein	

NRRL3_11763/NRRL3_11767 cluster	NRRL3_11765	An08g10880	fungal-specific transcription factor	
NRRL3_11763/NRRL3_11767 cluster	NRRL3_11766	An08g10870	2-methylcitrate dehydratase	
NRRL3_11763/NRRL3_11767 cluster	NRRL3_11767	An08g10860	fatty acid synthase beta subunit	AT-ER-DH-MPT
NRRL3_11763/NRRL3_11767 cluster	NRRL3_11768	An08g10830	polyprenyl synthetase family protein	
NRRL3_11763/NRRL3_11767 cluster	NRRL3_11769	An08g10820	aldehyde dehydrogenase	
Ochratoxin cluster	NFG	An15g07890	bAP-1 transcription factor	
Ochratoxin cluster	NFG	An15g07900	bcytochrome P450	
Ochratoxin cluster	NFG	An15g07910	bnon-ribosomal peptide synthetase	
Ochratoxin cluster	NRRL3_07207	An15g07920	polyketide synthase-like protein	[KS-AT-DH-MT-ER]-KR-(ACP)
Pyranonigrin E cluster	NRRL3_10119	An11g00350	FAD-dependent oxidoreductase PynB	
Pyranonigrin E cluster	NRRL3_10120	An11g00330	NAD(P)-binding protein PynE	
Pyranonigrin E cluster	NRRL3_10121	An11g00320	thioesterase domain-containing protein	
Pyranonigrin E cluster	NRRL3_10122	An11g00310	aspartic peptidase family protein	
Pyranonigrin E cluster	NRRL3_10123	An11g00300	FAD-binding domain-containing protein	
Pyranonigrin E cluster	NRRL3_10124[a]	An11g00290	fungal-specific transcription factor PynR	
Pyranonigrin E cluster	NRRL3_10125	An11g00280	O-methyltransferase PynC	
Pyranonigrin E cluster	NRRL3_10126	An11g00270	cytochrome P450 PynD	
Pyranonigrin E cluster	NRRL3_10127	An11g00260	MFS-type transporter PynF	
Pyranonigrin E cluster	NRRL3_10128	An11g00250	hybrid polyketide synthase/non-ribosomal peptide synthetase PynA	KS-AT-DH-KR-ER-KR-ACP--C-Rpt-A-T-[Te]
Siderophore cluster	NRRL3_08528	An03g03620	MFS-type transporter SitT	
Siderophore cluster	NRRL3_08529	An03g03600	helicase C-terminal domain-containing protein	
Siderophore cluster	NRRL3_08530	An03g03590	GNAT domain-containing protein	
Siderophore cluster	NRRL3_08531	An03g03580	hypothetical protein	
Siderophore cluster	NRRL3_08532	An03g03570	phospholipid methyltransferase-like protein	
Siderophore cluster	NRRL3_08533	An03g03570	hypothetical protein	
Siderophore cluster	NRRL3_08534	An03g03560	siderophore iron transporter	
Siderophore cluster	NRRL3_08535	An03g03550	crotonase superfamily protein	
Siderophore cluster	NRRL3_08536	An03g03540	hydroxyornithine transacylase SidF	
Siderophore cluster	NRRL3_08537	An03g03530	alpha/beta hydrolase fold domain-containing protein	
Siderophore cluster	NRRL3_08538	An03g03520	non-ribosomal peptide synthetase SidD	A-T-C-A-T-C
Siderophore cluster	NRRL3_08539	An03g03510	hypothetical protein	

Siderophore cluster	NRRL3_08540	An03g03500	tetratricopeptide-like helical domain-containing protein	
Siderophore cluster	NRRL3_11644	An06g01320	mevalonyl-CoA ligase SidI	
Siderophore cluster	NRRL3_11645	An06g01300	non-ribosomal peptide synthetase SidC	A-T-C-A-T-C-A-T-C-T-C-T-C
TAN-1612/BMS-192548 cluster	NRRL3_09545[a]	An11g07350	fungal-specific transcription factor AdaR	
TAN-1612/BMS-192548 cluster	NRRL3_09546[a]	An11g07340	O-methyltransferase AdaD	
TAN-1612/BMS-192548 cluster	NRRL3_09547[a]	An11g07330	FAD-dependent monooxygenase AdaC	
TAN-1612/BMS-192548 cluster	NRRL3_09548[a]	An11g07320	metallo-beta-lactamase-type thioesterase AdaB	
TAN-1612/BMS-192548 cluster	NRRL3_09549[a]	An11g07310	polyketide synthase AdaA	SAT-KS-AT-PT-ACP
TAN-1612/BMS-192548 cluster	NRRL3_09550	An11g07300	major facilitator superfamily protein	
Yanuthones cluster	NRRL3_06287[a]	An10g00100	fungal-specific transcription factor YanR	
Yanuthones cluster	NRRL3_06288	An10g00110	cytochrome P450 YanC	
Yanuthones cluster	NRRL3_06289	An10g00120	decarboxylase YanB	
Yanuthones cluster	NRRL3_06290	An10g00130	prenyltransferase YanG	
Yanuthones cluster	NRRL3_06291[a]	An10g00140	6-methylsalicylic acid synthase YanA	KS-AT-DH-KR-ACP
Yanuthones cluster	NRRL3_06292[a]	An10g00150	cytochrome P450 YanH	
Yanuthones cluster	NRRL3_06293[a]	An10g00160	O-mevalon transferase YanI	
Yanuthones cluster	NRRL3_06294	An10g00170	short-chain dehydrogenase/reductase YanD	
Yanuthones cluster	NRRL3_06295	An10g00180	RmlC-like cupin domain-containing protein YanE	
Yanuthones cluster	NRRL3_06296[a]	An10g00190	oxidase YanF	

[1]NRRL3_01579 and NRRL3_01580 share homology with FASs involved in primary metabolism and are not included in this table.
[2]Domains are written from the N-terminus to the C-terminus. Round brackets denote domains found in the NRRL 3 model but missing in the CBS 513.88 counterpart; domains unique to the CBS 513.88 model are in square brackets .A.: adenylation domain, ACP: acyl-carrier protein, AT: acyl transferase, C: condensation domain, DH: dehydratase, Dhe: dehydrogenase, ER: enoyl reductase, KR: ketoreductase, KS: ketosynthase, MPT: malonyl/palmitoyl transferase, MT: methyltransferase, NADB: NAD+ binding protein, PT: product template, PPT: phosphopantetheine transferase, Rpt: HxxPF repeat, SAT: starter unit acyltransferase, T: thiolation domain, Te: thioesterase/thioreductase.
[3]Genes found in only one *A. niger* strain are lebelled NFG (Not Found in Genome).
[4]Gene IDs labelled "not called" represent differences in annotated gene models between strains not DNA sequence.
[a]Gene/gene product has been experimentally characterized. The extent of the cluster is based on other means.
[b]Functional annotation of CBS 513.88 model assigned by curators.

A comparison of enzyme complement between *A. niger*, *A. fumigatus*, *A. oryzae*, and *A. nidulans* is shown in Table 5. A survey of backbone enzymes in these species reveals that the PKS/PKS-like and NRPS/NRPS-like genes make up the majority of all annotated backbone genes. The complement of NRPS and NRPS-like genes are also approximately equal in each of the four *Aspergilli*. Within the PKS category the majority of the genes encode the highly reducing class except in *A. fumigatus* and *A. oryzae* which have equivalent amounts of non-reducing and highly-reducing PKS genes. Partially reducing PKSs represent the smallest proportion of the PKS class in all species. The FAS complement is equal across the Aspergilli (5 or 6 FAS alpha and beta genes) with the exception of *A. fumigatus* which only has a single FAS alpha/beta gene pair. In the case of DMAT genes, the *A. niger* strains appear to have approximately half the number observed in *A. fumigatus*, *A. oryzae*, and *A. nidulans*. A number of backbone genes which we identified manually appear to have been missed by the SMURF and antiSMASH algorithms in the previous study annotating gene clusters in *A. niger* CBS 513.88 [27]. Looking at just the multi-domain backbone genes we were able to identify an additional PKS (An12g02050/ NRRL3_09034) and four additional NRPS-like genes (An01g06890/ NRRL3_02185, An01g11790/ NRRL3_02596, An01g14850/ NRRL3_02852 and An04g05420/ NRRL3_07812) in *A. niger* CBS513.88 along with their counter parts in the NRRL3 strain.

Table 5. Comparison of the total number of secondary metabolite backbone genes among four Aspergillus species and two strains of *A. niger*; NRRL3 and CBS513.88.

Backbone enzyme	***A. niger* strain NRRL3**	***A. niger* strain CBS 513.88**	***A. fumigatus***	***A. nidulans***	***A. oryzae***
FAS alpha/beta	5	5	1	6	5
HPN	8	7	1	1	2
HNP	1	1	0	0	0
NRPS	18	18	7	9	17
NRPS-like	21	21	6	13	13
Non reducing PKS	8	7	5	8	12
Highly reducing PKS	22	25	3	13	9
Partially reducing PKS	4	3	3	2	1
6-MSAS	1	1	0	0	1
PKS type III	1	1	1	2	5
PKS-like	4	4	1	1	0
Sesquiterpene cyclase	3	3	0	2	1
DMAT	2	2	5	7	6
Total	98	98	33	64	72

[1]The *A. nidulans* FASs comprised of 5 alpha subunits and 6 beta subunits.
[2]The *A. niger* strains contain one FAS alpha/beta (NRRL3_01579/NRRL3_01580) homologous to the FASs of primary metabolism.
[3]HNP enzyme represents a hybrid enzyme which contains an NRPS and PKS module in reverse order to HPNs

3.1.2 Experimentally characterized clusters with orthologues in *A. niger* NRRL3

A literature search for experimentally characterized SM clusters was carried out to look for experimentally verified clusters in other species which have orthologues in *A. niger* NRRL3 or in other *A. niger* strains which have counterparts in NRRL3. Our BLASTP queries revealed a total of 11 clusters in *A. niger* NRRL3 which have been assigned function based on experimentally determined secondary metabolite production in other fungal species and *A. niger* strains. The experimentally determined clusters were previously shown to produce: the azanigerones (*A. niger)* [100], TAN-1612 and its tautomer BMS-192548 (*A. niger*) [102], 1,8-

dihydroxynaphthalene (DHN) melanin and its pigment derivatives (in *A. niger* and *A. fumigatus*) [100, 105, 137], fumonisin (in *Gibberella moniliformis*) [122], ochratoxin (*A. niger*) [123], pyranonigrin (*A. niger*) [47], carlosic acid and agglomerin (*A. niger*) [99], the yanuthones (*A. niger*) [124], the kotanins (*A. niger*) [97], and two NRPS clusters with some genes shown to be involved in siderophore production (*A. fumigatus*) (Table 4) [138]. These experimentally defined clusters from other other fungal species and *A. niger* strains were nearly identical to the counterpart clusters in *A.niger* NRRL3 with respect to gene complement and arrangement. The exceptions were the fumonisin, DHN melanin and ochratoxin clusters. These clusters in *A. niger* NRRL3 either have their genes dispersed in the genome (DHN melanin) or lacked multiple genes previously annotated in other fungal species or *A. niger* strains (fumonisin and ochratoxin) [122, 123, 139] (Table 4). We also observed that one of these clusters, the azanigerone cluster, appears to be located within the FAS NRRL3_00135/NRRL3_00138/NRRL3_00156 cluster. The FAS cluster appears without the intervening azanigerone cluster in *A. acidus*. Conversely, the asperfuranone cluster which is nearly identical to the azanigerone cluster [100] appears without the FAS cluster in *A. nidulans* (Figure 14, Table 4).

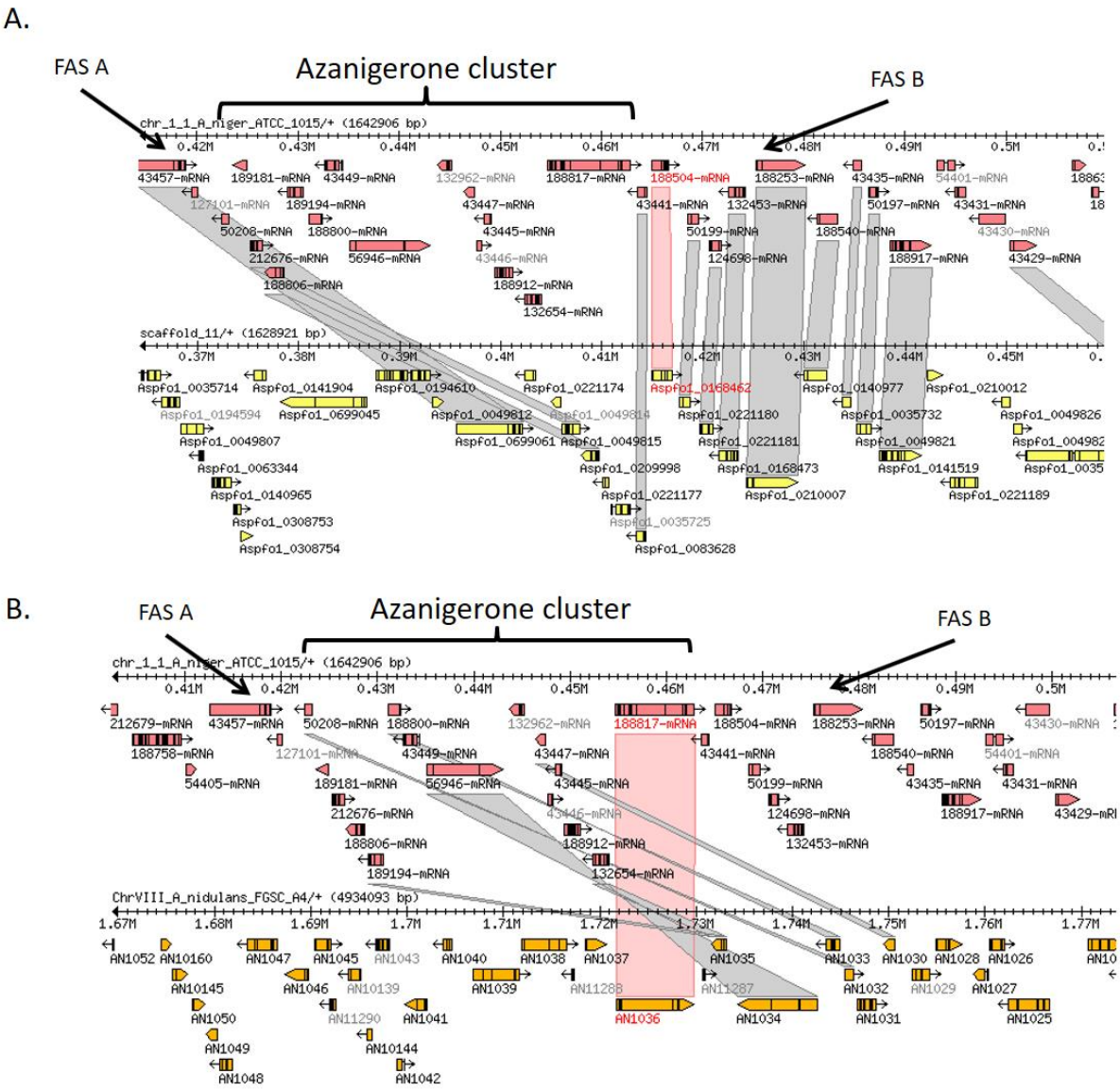


Figure 14. *A. niger* ATCC 1015 (top) synteny in *A. acidus* (bottom) showing sequence similarity between their FAS A/B clusters but not with the azanigerone cluster (A). *A. niger* ATCC 1015 (top) azanigerone cluster synteny with the similar asperfuranone cluster in *A. nidulans* (bottom) but not with the FAS A/B cluster of *A. niger* ATCC 1015 (B). Screenshot taken and from Sybil online application from aspGD website (See Materials and Methods).

3.1.3 *Aspergillus niger* NRRL3 and CBS 513.88: a genomic comparison of secondary metabolism between fungal strains

To examine differences between strains of *A. niger*, the annotated secondary metabolite backbone enzymes from the NRRL3 strain were compared to that of CBS 513.88 strain. The

comparison revealed differences with respect to enzyme absence/presence, enzyme truncation (remnants) and tailoring enzyme complement between clusters (Table 4, 6). Analysis of backbone enzymes revealed NRRL3 is missing three highly reducing PKS and one NRPS which are present in CBS 513.88. The reverse analysis shows one each of an NRPS, a partially reducing PKS and non-reducing PKS are present in NRRL3 but missing in CBS 513.88.

Table 6. Secondary metabolite cluster differences between *A. niger* strains NRRL3 and CBS 513.88.

	[1]NRRL3	[1]CBS 513.88
[2]Clusters specific to the NRRL3 lineage	NRRL3_08775 NRRL3_08790 NRRL3_10375	
[2]Clusters specific to the CBS 513.88 lineage		An01g01130 An11g05940 An11g05960 An15g07910
[3]Truncated clusters in NRRL3	NRRL3_04207 NRRL3_00166	[4]An15g07920 An09g02100
[3]Truncated clusters in CBS 513.88	NRRL3_03977 NRRL3_05440 NRRL3_00036 NRRL3_01367/01364 NRRL3_02593/02596 NRRL3_05440 NRRL3_06801 NRRL3_11726	An15g05090 An02g10140 An09g00520 An13g02460/02430 An01g11770/11790 An02g10140 An16g06720 An06g00430
CBS 513.88 clusters with truncated backbone enzymes	NRRL3_01804 NRRL3_09848	An01g02030 An11g03670

[1]Listed are genes encoding backbone enzymes.
[2]Cluster of genes encoding a full backbone enzyme plus one or more tailoring enzymes.
[3]Missing one or more genes encoding tailoring enzymes and may include a truncated backbone enzyme.
[4]Ochratoxin cluster.

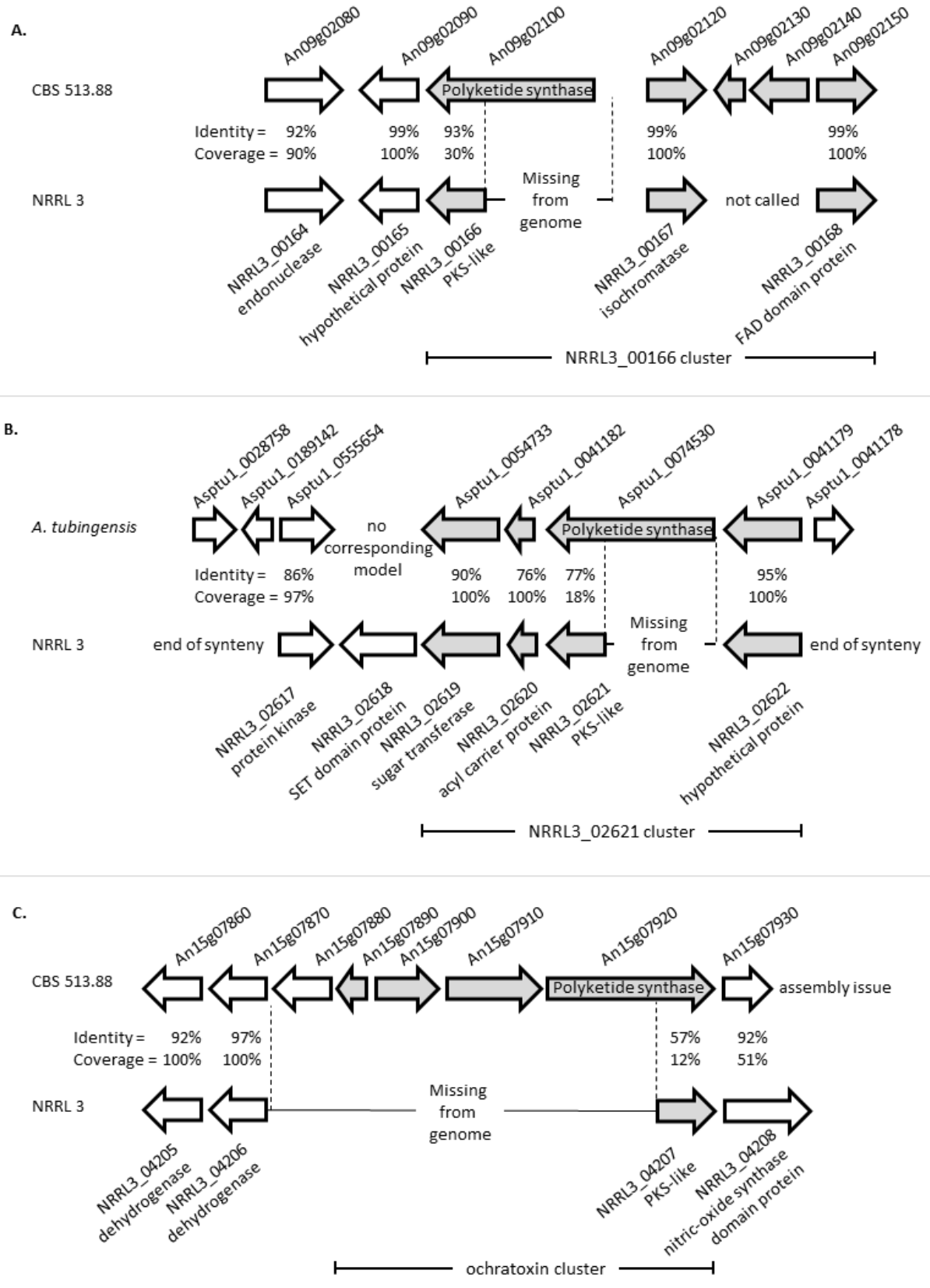
3.1.4 Backbone gene remnants in *A. niger* NRRL3 and CBS 513.88

Among the all the backbone genes found in the NRRL3 and CBS 513.88 genomes, we also located five remnant backbone genes. For simplicity only the remnant IDs are listed here;

the full PKS counterpart ID is omitted but can be found in Table 4 or Figure 15. Two remnants were unique to NRRL3 (NRRL3_00166, NRRL3_04207), two were unique to CBS 513.88 (An01g02030, An11g03670) and one remnant was common to both strains (NRRL3_02621/An01g12040). The unique remnants from one *A. niger* strain matched to (by BLASTP) to a complete counterpart in the opposite strain (NRRL3 vs CBS 513.88) while the shared remnant between both strains matched (by BLASTp) to a complete counterpart in *A. tubingensis* (Figure 15). All backbone remnant genes are exclusively of the PKS-like category and correspond to the 3' end of a complete PKS. The NRPS-like enzymes do not appear to have larger, multi-domain counterparts in other fungal species apart from NRRL3_02593/An01g11790 which had a counterpart in *A. brasiliensis* (Aspbr1_0027433) containing one additional C-terminal C domain.

The full PKS counterparts of the PKS-like enzymes are all highly reducing with the exception of NRRL3_00166 which corresponds to a PKS that contains an incongruous KR domain upstream of the KS domain (Table 4). Furthermore, the majority (four out of five) remnants contained at least the ACP minimal domain (NRRL3_00166, NRRL3_04207, An01g02030 and NRRL3_02621/An01g12040). Lastly, a cluster analysis of remnant enzymes revealed that in three out of five remnants only the backbone enzyme is missing while surrounding genes are retained (Figure 15 A, B, E). In the other two remnant clusters, NRRL3_04207 and An01g02030, additional surrounding genes are missing that are present in the opposing *A. niger* strain containing the full highly reducing PKS (Figure 15 C and D).

An additional search to find more remnant backbone genes in the NRRL3 strain was carried out using the peptide sequences of 114 genes containing single SM domains. These genes were not identified by the initial backbone search however their single domains were annotated from InterProScan (manuscript in preparation) as belonging to one of the SM backbone multi-domain enzymes (PKS, NRPS or FAS). Surmising that the single domain enzymes may themselves be remnants, I queried these enzymes by BLASTP against other fungal genomes from the AspGD website. The query consisted of sequences from the following domains: 1 PKS-KS, 1 PKS-AT, 20 PKS-KR, 36 PKS-ER, 2 PKS-ACP, 33 NRPS-A, 2 FAS-PPT, and 17 FAS-ACP. None of the single domain enzymes appeared to have a full PKS, NRPS or FAS counterpart in other fungal species.


A.
An09g02080
An09g02090
An09g02100
An09g02120
An09g02130
An09g02140
An09g02150
CBS 513.88
Polyketide synthase
Identity = 92% 99% 93% 99% 99%
Coverage = 90% 100% 30% 100% 100%
NRRL 3
Missing from genome
not called
NRRL3_00164 endonuclease
NRRL3_00165 hypothetical protein
NRRL3_00166 PKS-like
NRRL3_00167 isochromatase
NRRL3_00168 FAD domain protein
NRRL3_00166 cluster
B.
Asptu1_0028758
Asptu1_0189142
Asptu1_0555654
Asptu1_0054733
Asptu1_0041182
Asptu1_0074530
Asptu1_0041179
Asptu1_0041178
A. tubingensis
no corresponding model
Polyketide synthase
Identity = 86% 90% 76% 77% 95%
Coverage = 97% 100% 100% 18% 100%
NRRL 3
end of synteny
Missing from genome
end of synteny
NRRL3_02617 protein kinase
NRRL3_02618 SET domain protein
NRRL3_02619 sugar transferase
NRRL3_02620 acyl carrier protein
NRRL3_02621 PKS-like
NRRL3_02622 hypothetical protein
NRRL3_02621 cluster
C.
An15g07860
An15g07870
An15g07880
An15g07890
An15g07900
An15g07910
An15g07920
An15g07930
CBS 513.88
Polyketide synthase
assembly issue
Identity = 92% 97% 57% 92%
Coverage = 100% 100% 12% 51%
NRRL 3
Missing from genome
NRRL3_04205 dehydrogenase
NRRL3_04206 dehydrogenase
NRRL3_04207 PKS-like
NRRL3_04208 nitric-oxide synthase domain protein
ochratoxin cluster

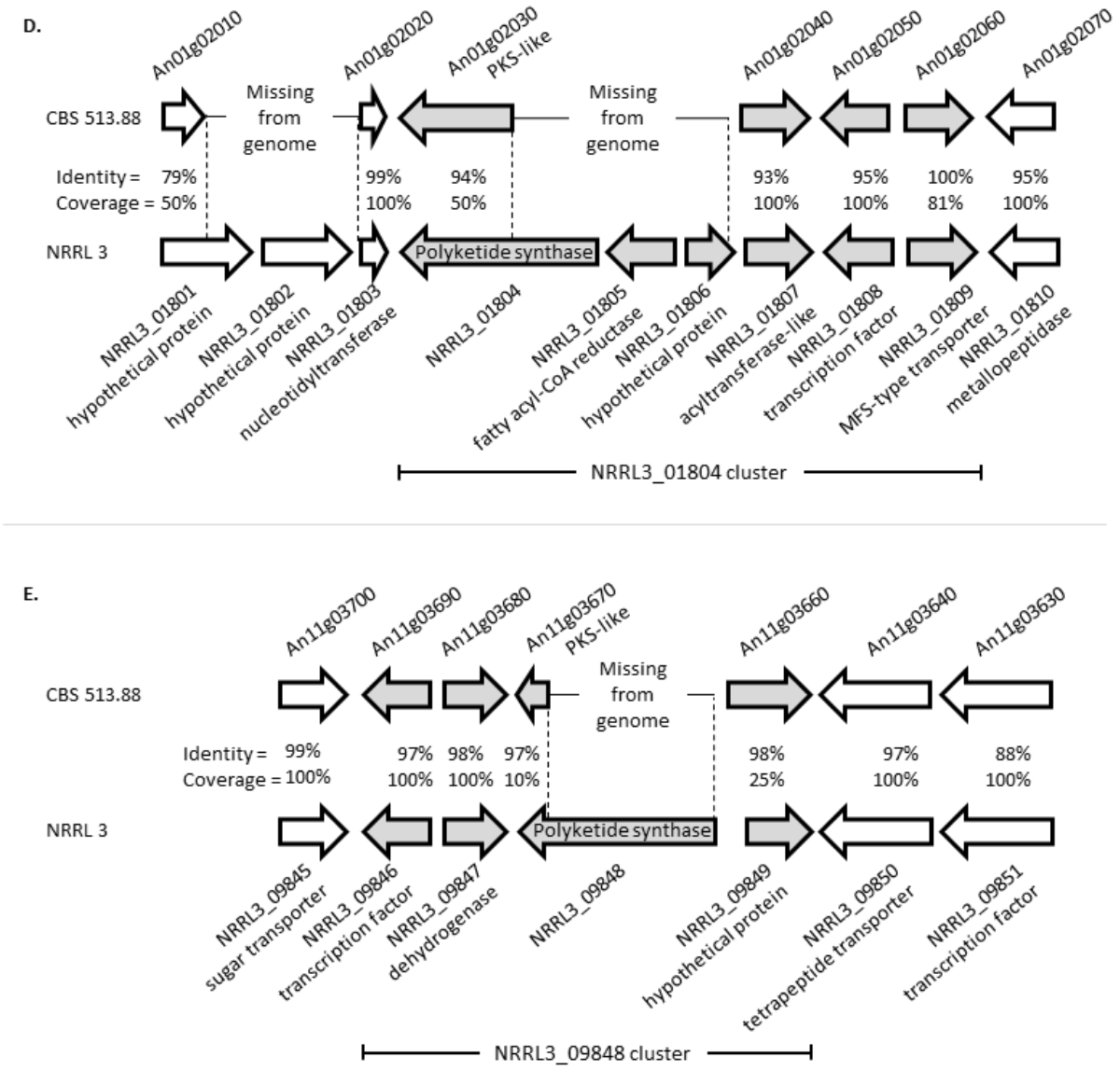


Figure 15. Remnant PKS-like enzymes from *A. niger* NRRL3 and CBS 513.88 aligned with their full enzyme counterparts.

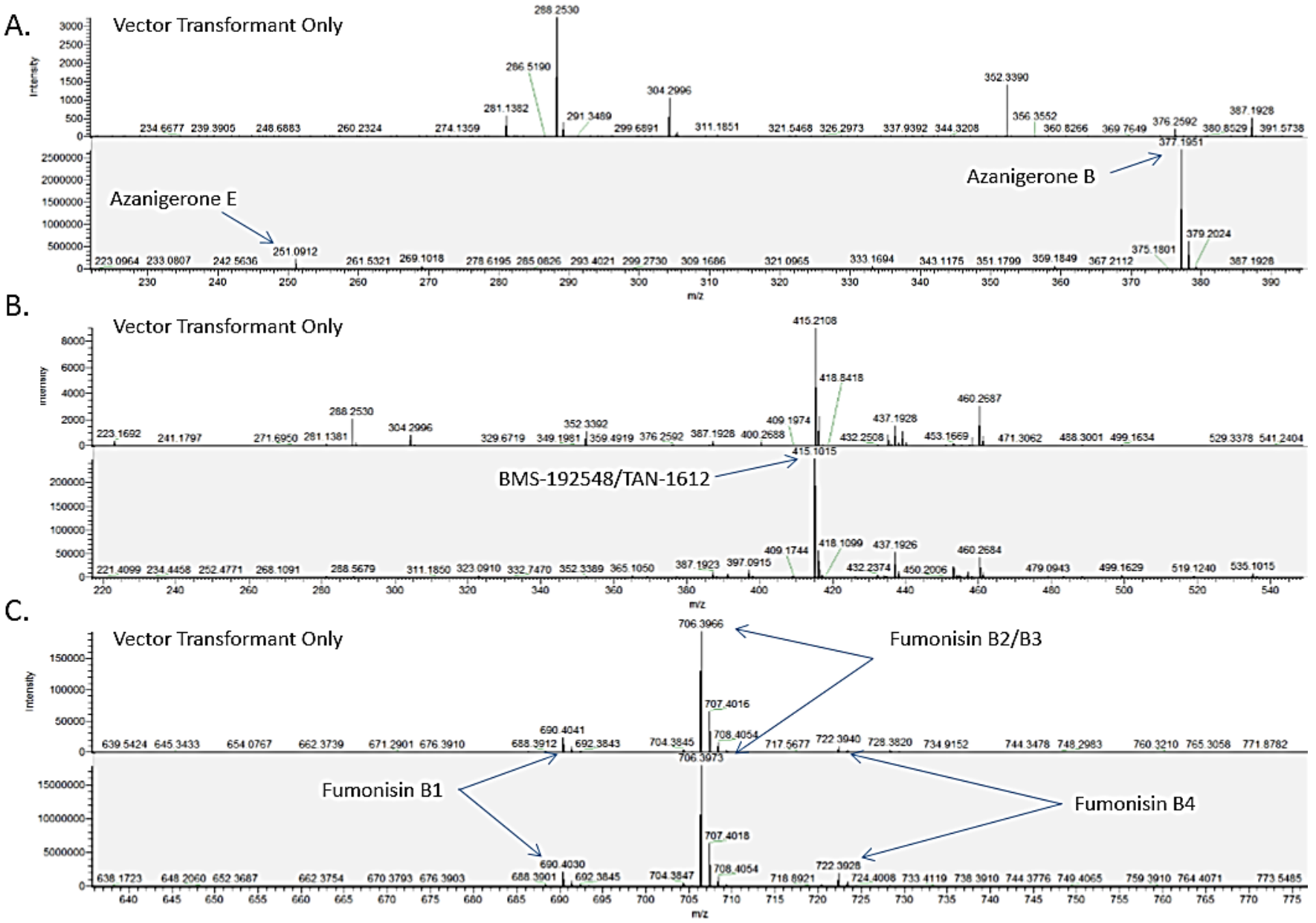
3.2 Upregulation of defined clusters

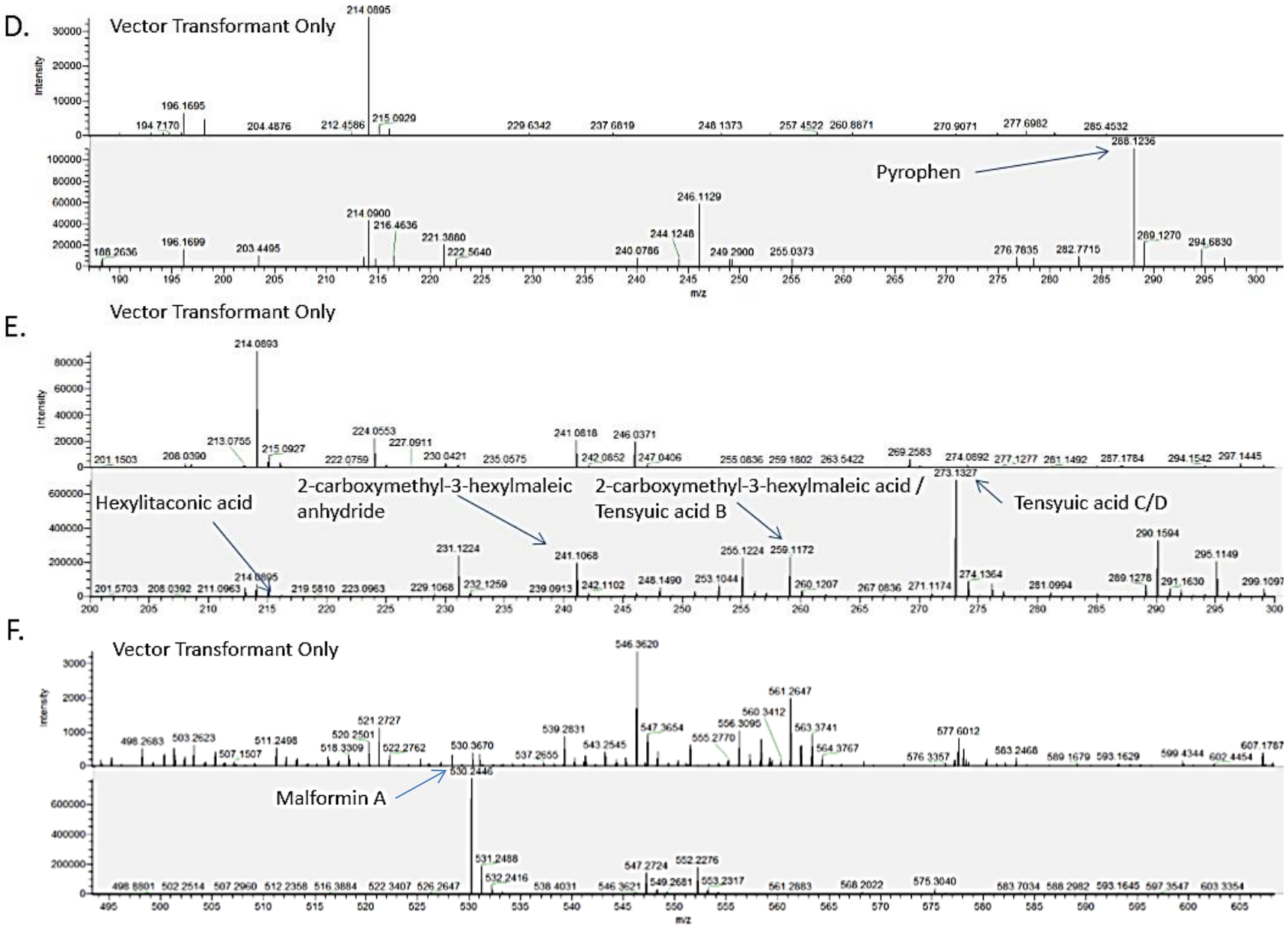
3.2.1 Expression of genes encoding clustered transcription factors

Given that 43 of the 84 SM clusters appear to have at least one resident transcription factor, we sought to overexpress some of these transcription factors in *A. niger* in an effort to induce cluster expression. We selected 35 transcription factors within or just outside 31 clusters

for overexpression. These included the transcription factors of three known SM clusters which we used as controls: fumonisin (NRRL3_02186) [122], BMS-192548/TAN-1612 (NRRL3_09545) [102], and the azanigerones (NRRL3_00148) [100]. Both BMS-192548/TAN-1612 and the azanigerones are also colored compounds which allowed for quick phenotypic assessment of SM production following growth of transformants on selective media [100, 102].

Of all the overexpressed transcription factors, six (including the controls) stimulated the production of identifiable compounds (NRRL3_00148, NRRL3_00406, NRRL3_02186, NRRL3_07873, NRRL3_09545, and NRRL3_11765) and one (NRRL3_08781) appeared to stimulate the production of compounds which we could not identify (Figure 16). The regulators of the malformins (NRRL3_07873), pyrophen (NRRL3_00406) and tensyuic acids/hexylitaconic acid (NRRL3_11765) (herein referred to cumulatively as alkyl citric acids) are to our knowledge currently unknown and potentially constitute a novel means of overexpression for these orphan compounds. While the identity of the compounds generated from the overexpression of NRRL3_08781 is unknown, the transcription factor itself is situated between a PKS and an NRPS cluster. This indicates that the compounds may be one of set of polyketides, short peptides or a hybrid if the PKS and NRPS are within the same cluster.


A.
Vector Transformant Only
Azanigerone E
Azanigerone B
B.
Vector Transformant Only
BMS-192548/TAN-1612
C.
Vector Transformant Only
Fumonisin B2/B3
Fumonisin B1
Fumonisin B4
Intensity
m/z


D.
Vector Transformant Only
Intensity
214.0895
196.1695
194.7170
204.4876
212.4586
215.0929
229.6342
237.6619
248.1373
257.4522
260.8871
270.9071
277.6982
285.4532
288.1236
Pyrophen
246.1129
214.0900
216.4636
221.3660
222.5640
244.1248
240.0786
249.2900
255.0373
276.7635
282.7715
289.1270
294.6830
186.2636
196.1699
203.4495
m/z
E.
Vector Transformant Only
214.0893
213.0755
215.0927
224.0553
227.0911
230.0421
235.0575
241.0618
246.0371
242.0852
247.0406
255.0836
259.1802
263.5422
269.2583
274.0892
277.1277
281.1492
287.1784
294.1542
297.1445
201.1503
208.0390
222.0759
273.1327
Hexylitaconic acid
2-carboxymethyl-3-hexylmaleic anhydride
2-carboxymethyl-3-hexylmaleic acid / Tensyuic acid B
Tensyuic acid C/D
231.1224
241.1068
255.1224
259.1172
290.1594
295.1149
214.0895
201.5703
208.0392
211.0963
219.5810
223.0963
229.1068
232.1259
239.0913
242.1102
248.1490
253.1044
260.1207
267.0836
271.1174
274.1364
281.0994
289.1278
291.1630
299.1097
F.
Vector Transformant Only
Intensity
546.3620
561.2647
560.3412
556.3095
563.3741
577.6012
521.2727
520.2501
539.2831
547.3654
555.2770
564.3767
498.2683
503.2623
511.2498
507.1507
518.3309
522.2762
530.3670
537.2655
543.2545
576.3357
583.2468
589.1679
593.1629
599.4344
602.4454
607.1787
530.2446
Malformin A
531.2488
547.2724
552.2276
532.2416
538.4031
546.3621
549.2681
553.2317
561.2683
568.2022
575.3040
498.8801
502.2514
507.2960
512.2358
516.3884
522.3407
526.2647
583.7034
588.2982
593.1645
597.3547
603.3354
m/z

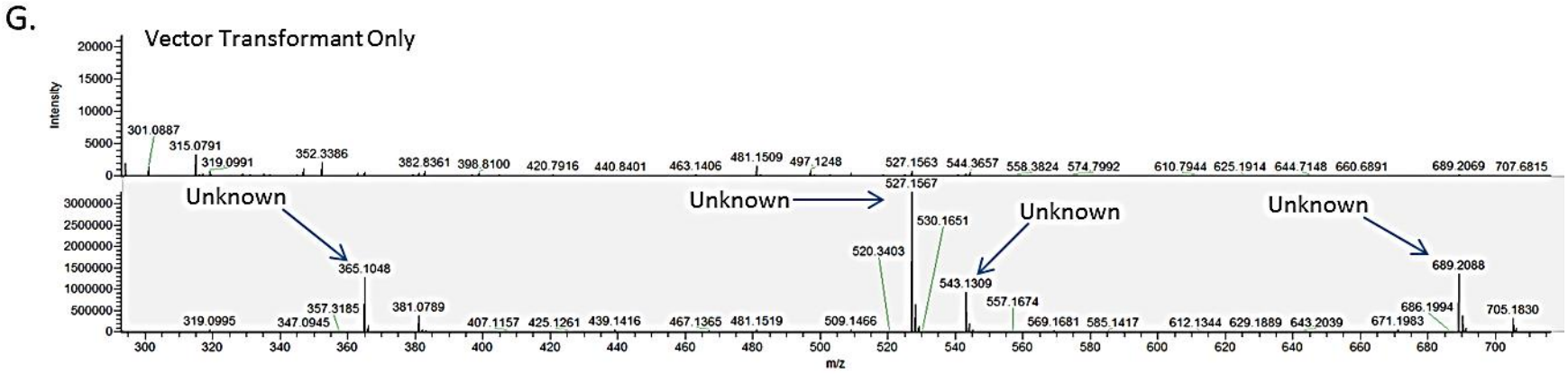


Figure 16. Mass spectra of transcription factor overexpression strains overproducing identifiable compounds. The control strain is the vector transformant only (ANIp7) (A) NRRL3_00148 overexpression produces the azanigerones. (B) NRRL3_09545 overexpression produces BMS-192548/TAN-1612. (C) NRRL3_02186 overexpression increases production of fumonisins compared to controls. (D) NRRL3_00406 overexpression produces pyrophen. (E) NRRL3_11765 overexpression produces the tensyuic acids and other alkyl citric acids. (F) NRRL3_07873 overexpression produces malformin A. (G) NRRL3_08781 generates a set of unknown peaks. The 214.0892 m/z observed in the positive spectra of the NRRL3_11765 overexpression and control strains correspond to the mass of n-butyl benzene sulfonamide; a common mass spectrometry contaminant [140].

Among the transcription factors which generated identifiable (by mass spectrometry) compounds, our focus shifted to a single transcription factor (NRRL3_11765) where the mass spectrometry (MS) data revealed the presence the tensyuic acids (Figure 16E). We selected these compounds for pathway reconstruction since, to our knowledge, they have no known genetic underpinnings and had been previously shown to possess useful biological properties [93, 94]. Three other alkyl citric acid masses corresponding to 2-carboxymethyl-3-hexylmaleic acid, its anhydride and hexylitaconic acid were also identified (Figure 16E).

3.2.2 Overexpression of transcription factor NRRL3_11765: transcriptome analysis

Examining the upregulated genes in the RNA-seq transcriptome of the NRRL3_11765 overexpression strain (Table 7), we could determine which genes are likely involved in the biosynthesis of the alkyl citric acids. Based on these expression we refined the previously annotated cluster which now included a gene of unknown function (NRRL3_11757) and excluded a transcription factor (NRRL3_11760), a zinc finger protein (NRRL3_11762) and a polyprenyl synthase (NRRL3_11768) (Figure 17). The up-regulated genes in the cluster span from gene NRRL3_11757 to NRRL3_11769. Based on the functional annotation of the up-regulated genes within the cluster, we could only account for the production of the 2-carboxymethyl-3-hexylmaleic acid compound. We therefore assigned a *cha* (for 2-carboxymethyl-3-hexylmaleic acid) identifier to the genes that are up-regulated in the NRRL3_11765 overexpression strain in this region. We therefore refer to the NRRL3_11765 transcription factor as *chaR*, the overexpressing strain as $chaR^{OE}$ and the parent strain as the control strain. We also noted that another cluster containing genes (NRRL3_02448 – NRRL3_02450) with predicted functions to generate itaconic acid from citric acid [141] is down-regulated in $chaR^{OE}$ (Table 8). Itaconic acid was not detected in the extracellular growth medium or intracellular fractions of any of our strains (data not shown).

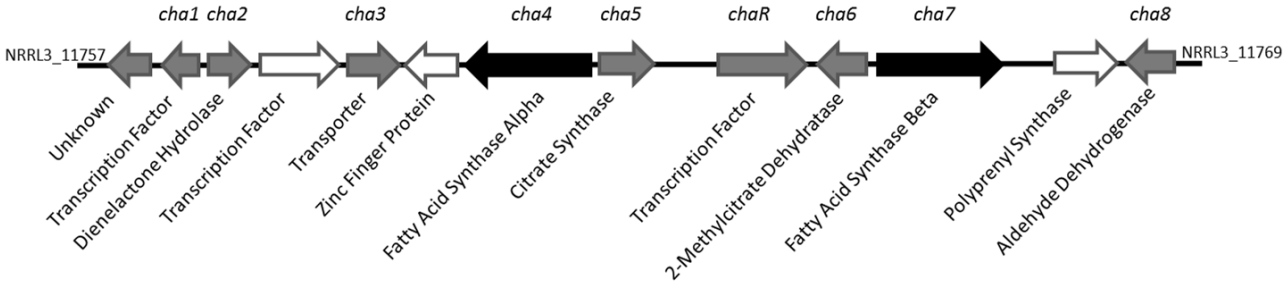


Figure 17. Clustered genes upregulated (grey and black arrows) by *chaR* and clustered genes whose expression is unaffected (white arrows) by overexpression of *chaR*. Labelled *cha* genes represent backbone (black arrows) and accessory enzymes (grey arrows) determined to be involved in the production of hexylitaconic acid.

Table 7. Genes up regulated in the *chaR*^{OE} strain compared to controls. Highlighted portion indicates cha cluster.

Gene ID	FPKM control strain	FPKM *chaR*	log2 Fold Change	p-value	Predicted function
NRRL3_11769	2.82	19.2	2.77	1.69E-05	Aldehyde dehydrogenase
NRRL3_11767	0.06	36.1	9.23	8.94E-30	Fatty acid synthase subunit beta
NRRL3_11766	0.06	82.98	10.43	1.72E-29	2-methylcitrate dehydratase
NRRL3_11765	0.02	181.95	13.15	9.09E-41	Fungal specific transcription factor
NRRL3_11764	0.01	1892.31	17.53	1.51E-55	Citrate synthase
NRRL3_11763	0.09	1241.98	13.75	3.41E-53	Fatty-acid synthase
NRRL3_11761	6.39	852.5	7.06	1.61E-24	MFS-type transporter
NRRL3_11759	22.35	1547.48	6.11	6.85E-20	Dienelactone hydrolase family protein
NRRL3_11758	2.19	23.51	3.42	1.71E-04	Fungal specific transcription factor
NRRL3_11757	5.9	62.65	3.41	2.20E-07	Unknown function
NRRL3_11501	2.15	17.03	2.99	2.14E-05	Amidohydrolase 2 family protein
NRRL3_11500	88.84	1262.62	3.83	3.67E-11	Aldo/keto reductase family protein
NRRL3_11100	3.53	23.39	2.73	3.00E-05	Short-chain dehydrogenase
NRRL3_10103	16.72	687.64	5.36	4.98E-17	ATP-citrate synthase subunit 1
NRRL3_10102	17.7	593.56	5.07	6.70E-16	ATP-citrate synthase subunit
NRRL3_09843	2.19	51.75	4.56	1.05E-12	FAD-linked oxioreductase family protein
NRRL3_09788	1.5	139.39	6.54	1.39E-18	Short-chain dehydrogenase
NRRL3_09757	1.41	22.31	3.98	1.18E-08	Aldo/keto reductase family protein
NRRL3_08741	2.51	15.45	2.62	3.65E-05	Unknown function
NRRL3_08622	15.74	84.03	2.42	8.34E-06	Short-chain dehydrogenase
NRRL3_08414	12.72	59.69	2.23	2.64E-05	Serine hydrolase FSH1 family protein
NRRL3_08394	1.17	14.84	3.66	1.54E-06	Fungal specific transcription factor
NRRL3_08390	4.09	24.57	2.59	8.50E-05	NmrA-like family protein
NRRL3_08383	0	152.45	Inf	6.66E-37	O-methyltransferase family protein
NRRL3_08061	19.2	93.78	2.29	2.07E-05	L-Serine ammonia-lyase
NRRL3_07921	4.09	40.71	3.32	3.72E-06	Short-chain dehydrogenase
NRRL3_07736	7.35	256.7	5.13	2.64E-15	Fungal specific transcription factor
NRRL3_07496	2.33	34.4	3.88	7.29E-10	Transmembrane amino acid transporter protein
NRRL3_07364	1.62	15.07	3.22	1.30E-06	MFS-type transporter
NRRL3_07196	2.12	28.96	3.77	1.76E-09	Reverse transcriptase / Endonuclease
NRRL3_07007	4.78	26.28	2.46	3.60E-05	C6-zinc finger domain-containing protein
NRRL3_06970	8.42	90.73	3.43	3.51E-09	AMP-dependent synthetase
NRRL3_06778	9.53	57.11	2.58	2.75E-05	3-oxoacyl-[acyl-carrier-protein] reductase
NRRL3_06474	16.06	99.24	2.63	2.05E-06	Short-chain dehydrogenase
NRRL3_06447	3	38.38	3.68	3.37E-09	Zinc-type alcohol dehydrogenase
NRRL3_06436	1.02	63.29	5.96	4.87E-17	Beta-glucosidase
NRRL3_06423	92.81	804.03	3.11	1.70E-08	Aldo/keto reductase family protein

NRRL3_06323	5.93	41.95	2.82	4.45E-06	Endonuclease III-like protein
NRRL3_06281	0.97	17.51	4.17	1.48E-07	Zinc-type alcohol dehydrogenase
NRRL3_06279	1.16	18.65	4.01	9.77E-09	MFS-type transporter
NRRL3_06229	0.58	12.38	4.42	1.47E-09	Fungal specific transcription factor
NRRL3_06131	21.24	131.9	2.63	2.80E-06	NADPH-adrenodoxin reductase, mitochondrial
NRRL3_06061	0.44	23.53	5.74	5.59E-11	Unknown function
NRRL3_05609	9.52	59.48	2.64	1.70E-05	NAD(P)-binding domain-containing protein
NRRL3_05463	44.53	185.87	2.06	1.60E-04	GNS1/SUR4 membrane family protein
NRRL3_05398	5.98	70.35	3.56	5.85E-08	Short-chain dehydrogenase
NRRL3_04990	0.48	11.66	4.60	2.22E-08	Beta-glucosidase
NRRL3_04324	4.94	54.14	3.45	3.94E-09	ABC-2 type transporter
NRRL3_04323	5.55	99.76	4.17	4.82E-11	MFS-type transporter
NRRL3_03973	1.85	69.79	5.24	3.01E-14	Disease resistance family protein
NRRL3_03750	1.34	41.46	4.95	9.16E-12	FAD- domain protein/monooxygenase
NRRL3_03531	43.04	179.24	2.06	8.20E-05	Cystathionine beta-synthase
NRRL3_03431	61.4	452.81	2.88	2.84E-06	Acetyl-CoA carboxylase (ACC)
NRRL3_03003	8.67	59.52	2.78	2.91E-06	Cytochrome P450 family protein
NRRL3_02974	1.75	39.78	4.51	6.07E-10	Acyl-CoA desaturase
NRRL3_02906	2.22	39.14	4.14	1.87E-09	NAD(P)-binding domain-containing protein
NRRL3_02295	3	56.22	4.23	7.53E-10	Short-chain dehydrogenase
NRRL3_02207	25.14	153.34	2.61	3.50E-06	Salicylate hydroxylase
NRRL3_02206	5.49	35.93	2.71	3.99E-06	Hydantoinase B/oxoprolinase family protein
NRRL3_02101	12.82	63.15	2.30	3.08E-05	Fungal specific transcription factor
NRRL3_02100	4.76	115.48	4.60	1.05E-12	MFS-type transporter
NRRL3_01768	59.17	497.8	3.07	3.32E-08	NAD dependent epimerase
NRRL3_01429	1.77	15.98	3.17	8.17E-06	Short-chain dehydrogenase
NRRL3_01224	12.1	52.92	2.13	8.14E-05	Carbohydrate-Binding Module Family 18
NRRL3_01223	20.28	78.58	1.95	1.45E-04	Glycoside Hydrolase Family 55
NRRL3_01212	8.35	60.4	2.85	6.10E-07	Carbohydrate-Binding Module Family 50
NRRL3_01029	2.98	65.31	4.45	2.60E-10	Short-chain dehydrogenase
NRRL3_00829	4.47	42.51	3.25	2.41E-07	Conserved unknown protein
NRRL3_00745	2.66	242.79	6.51	3.77E-19	Unknown function
NRRL3_00700	2.27	21.72	3.26	1.01E-05	Short-chain dehydrogenase
NRRL3_00546	191.64	1145.5	2.58	3.62E-06	Mitochondrial citrate/oxoglutarate carrier protein
NRRL3_00504	2.07	13.45	2.70	1.70E-04	Cis-aconitate decarboxylase
NRRL3_00386	36.48	169	2.21	1.26E-05	MFS-type Sugar/inositol transporter
NRRL3_00357	3.63	23.42	2.69	3.44E-05	Conserved unknown protein
NRRL3_00308	4.19	81.05	4.27	1.86E-06	Unknown function

*Inf = Infinite

Table 8. Genes down-regulated in the *chaR*^{OE} strain compared to controls. Highlighted portion indicates putative itaconate gene cluster.

Gene ID	FPKM control strain	FPKM *chaR*	log2 Fold Change	p-value	Predicted function
NRRL3_11794	49.52	1.4	-3.25	2.24E-09	Conserved unknown protein
NRRL3_11673	24.63	1.06	-2.86	3.56E-07	Basic-leucine zipper (bZIP) transcription factor
NRRL3_11598	47.75	7.79	-1.65	1.60E-04	Amidase, hydantoinase
NRRL3_11498	304.18	27.9	-2.17	1.02E-06	Cupin superfamily protein
NRRL3_11493	31.46	2.87	-2.18	1.60E-05	Unknown function
NRRL3_11028	182.51	26.6	-1.75	5.69E-05	SnoaL-like domain-containing protein
NRRL3_10386	27.95	0.77	-3.27	6.03E-08	Unknown function
NRRL3_10374	65.56	8.35	-1.88	9.33E-05	Zinc-type alcohol dehydrogenase
NRRL3_10371	29.86	3.11	-2.06	1.78E-04	Serine hydrolase FSH1 family protein
NRRL3_10369	66.51	7.31	-2.01	2.45E-05	Methyltransferase type 11 domain-containing protein
NRRL3_10215	197.8	28.26	-1.77	1.39E-04	Heat shock protein 20
NRRL3_10109	60.74	5.62	-2.17	1.38E-06	MFS-type transporter
NRRL3_09550	131.31	1.05	-4.40	5.90E-21	MFS-type transporter
NRRL3_09477	46.36	4.37	-2.15	1.52E-06	Cytochrome P450 family protein
NRRL3_09010	101.68	14.55	-1.77	6.54E-05	Amidohydrolase family protein
NRRL3_08986	33.71	2.58	-2.34	7.12E-06	Methyltransferase domain-containing protein
NRRL3_08850	31.73	3.12	-2.11	1.70E-05	Amidase family protein
NRRL3_08747	44.95	5.13	-1.98	7.41E-05	Heat shock protein 20
NRRL3_08658	1556.54	239.27	-1.70	9.12E-05	Unknown function
NRRL3_08340	135.51	1.2	-4.30	3.34E-13	Carboxymuconolactone decarboxylase family protein
NRRL3_08288	97.09	5.48	-2.62	1.53E-09	AAA ATPase domain-containing protein
NRRL3_07887	235.88	25.15	-2.04	2.58E-05	Unknown function
NRRL3_07886	24.79	4.18	-1.62	1.42E-04	Tetratricopeptide-like helical domain-containing protein
NRRL3_07772	40.23	5.22	-1.86	2.33E-05	Linoleate 8R-lipoxygenase
NRRL3_07337	11.78	0.62	-2.68	2.50E-05	Unknown function
NRRL3_06897	59.59	5.68	-2.14	6.65E-06	Cytochrome P450 family protein
NRRL3_06872	25.22	5.05	-1.46	1.42E-04	Tetratricopeptide-like helical domain-containing protein
NRRL3_06728	136.95	17.53	-1.87	3.69E-05	C2H2-like zinc finger domain-containing protein
NRRL3_06394	223.43	18.58	-2.26	1.18E-06	EBP family protein
NRRL3_06331	71.98	5.89	-2.28	1.17E-06	1-aminocyclopropane-1-carboxylate deaminase
NRRL3_06275	41.31	2.04	-2.74	1.04E-10	Unknown function
NRRL3_06189	46.8	6.75	-1.76	3.10E-05	Nonribosomal peptide synthetase-like enzyme
NRRL3_05997	121.4	19.18	-1.68	8.57E-05	3-isopropylmalate dehydratase
NRRL3_05625	14.37	0.96	-2.46	3.27E-07	Linoleate 10R-lipoxygenase
NRRL3_05588	14.57	2.63	-1.56	4.74E-05	Hybrid PKS/NRPS
NRRL3_05587	1052.45	106.59	-2.08	3.29E-06	EthD domain-containing protein
NRRL3_05586	61.8	5.45	-2.21	6.36E-06	Zinc-type alcohol dehydrogenase

NRRL3_05464	19.84	0.08	-5.02	3.76E-06	Unknown function
NRRL3_04967	45.12	3.9	-2.23	4.51E-05	Unknown function
NRRL3_04660	308.04	14.58	-2.78	1.30E-10	Unknown function
NRRL3_04562	16.84	2.21	-1.85	3.68E-05	HET domain-containing protein
NRRL3_04536	14.83	0.84	-2.61	2.49E-06	Amidase
NRRL3_04409	365.88	48.37	-1.84	2.55E-05	Glutathione S-transferase
NRRL3_04122	287.64	24.5	-2.24	1.52E-07	NADH-cytochrome b5 reductase
NRRL3_03871	168.99	2.25	-3.93	3.51E-13	Fungal hydrophobin family protein
NRRL3_03740	22.64	1.45	-2.50	7.73E-06	Unknown function
NRRL3_03591	168.78	12.09	-2.40	4.72E-07	Polysaccharide deacetylase
NRRL3_03284	10.81	0.68	-2.52	1.23E-07	Transferase family protein
NRRL3_03199	55.9	6.46	-1.96	2.24E-05	Metallo-dependent hydrolase
NRRL3_03198	243.32	37.67	-1.70	1.38E-04	Guanyl-specific ribonuclease
NRRL3_03130	200.84	23.89	-1.94	8.94E-06	Lactam utilization protein lamB
NRRL3_02998	356.24	57.14	-1.67	1.33E-04	Nitrite transporter, high-affinity
NRRL3_02996	95.28	7.43	-2.32	2.36E-07	GTP cyclohydrolase II family protein
NRRL3_02994	40.22	3.64	-2.19	2.25E-05	Uracil phosphoribosyltransferase
NRRL3_02986	178.86	22.14	-1.90	1.56E-05	Alpha/beta hydrolase fold domain-containing protein
NRRL3_02951	11.17	0.74	-2.47	2.86E-06	MFS-type transporter
NRRL3_02862	46.39	4.67	-2.09	2.28E-05	L-asparaginase
NRRL3_02821	27.75	1.71	-2.54	5.32E-07	FAD-linked oxioreductase family protein
NRRL3_02561	94.3	6.1	-2.49	1.27E-08	NADH:flavin oxidoreductase
NRRL3_02560	399.35	22.49	-2.62	1.99E-09	NmrA-like family protein
NRRL3_02557	218.09	18.01	-2.27	3.13E-07	Methionine aminopeptidase 1
NRRL3_02534	40.8	2.21	-2.65	2.62E-08	NmrA-like family protein
NRRL3_02450	33.78	1.25	-3.00	2.20E-09	Cis-aconitate decarboxylase
NRRL3_02449	18.1	2.15	-1.94	1.31E-04	Citrate synthase
NRRL3_02448	12.08	0.18	-3.83	2.11E-07	2-methylcitrate dehydratase
NRRL3_01961	18.03	0.78	-2.86	1.38E-09	ABC transporter
NRRL3_01724	153.1	4.51	-3.21	6.67E-12	Alpha/beta hydrolase fold domain-containing protein
NRRL3_01672	46.37	3.67	-2.31	1.98E-06	FAD dependent oxidoreductase domain-containing protein
NRRL3_01443	63.78	0.36	-4.71	7.30E-09	Unknown function
NRRL3_01338	32.17	2.12	-2.48	3.63E-08	Calycin domain-containing protein
NRRL3_01334	11.93	1.19	-2.10	2.34E-06	Non-Linear Nonribosomal peptide synthetase
NRRL3_01127	64.17	5.72	-2.20	1.31E-06	Dihydroxyacetone kinase
NRRL3_00992	16.3	0.1	-4.64	2.58E-10	Fatty acyl-CoA reductase
NRRL3_00733	11.46	0.5	-2.85	1.99E-06	Fungal specific transcription factor
NRRL3_00479	40	4.85	-1.92	1.94E-05	Sulfhydryl oxidase
NRRL3_00287	468.88	14.52	-3.16	4.05E-11	Glutathione-dependent formaldehyde-activating family protein
NRRL3_00247	16.63	0.53	-3.14	8.76E-07	Carbonic anhydrase
NRRL3_00177	16.88	1.62	-2.13	1.08E-05	Carboxylesterase, type B family protein
NRRL3_00095	13.82	1.43	-2.06	4.74E-05	Tetratricopeptide-like helical domain-containing protein
NRRL3_00027	46.73	4.43	-2.14	1.63E-05	HAD-like hydrolase superfamily protein

3.2.3 Overexpression of transcription factor NRRL3_11765: Compound identification by nuclear magnetic resonance (NMR)

We used NMR to elucidate the structures of compounds produced by the *chaR*^{OE} strain. The NMR analysis of isolated compounds from media extracts of the *chaR*^{OE} strain identified the structures of hexylitaconic acid, 2-carboxymethyl-3-hexylmaleic acid, hexylcitraconic acid, a hydroxylated form of hexylitaconic acid and a carbonylated form of hexylitaconic acid (Figure 18, Table 9). No structural solutions were obtained for the anhydride form of 2-carboxymethyl-3-hexylmaleic acid, a compound that has previously been identified from *A. niger* [92], or any of the tensyuic acids and they remain unconfirmed and putatively identifiable only by MS.

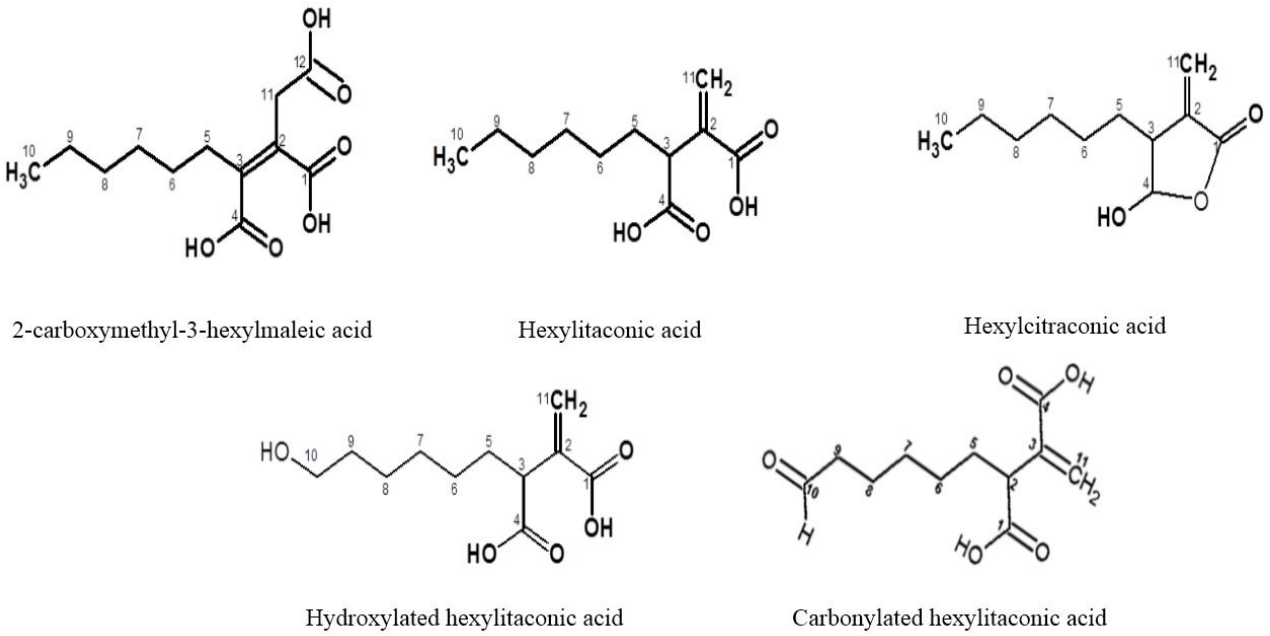


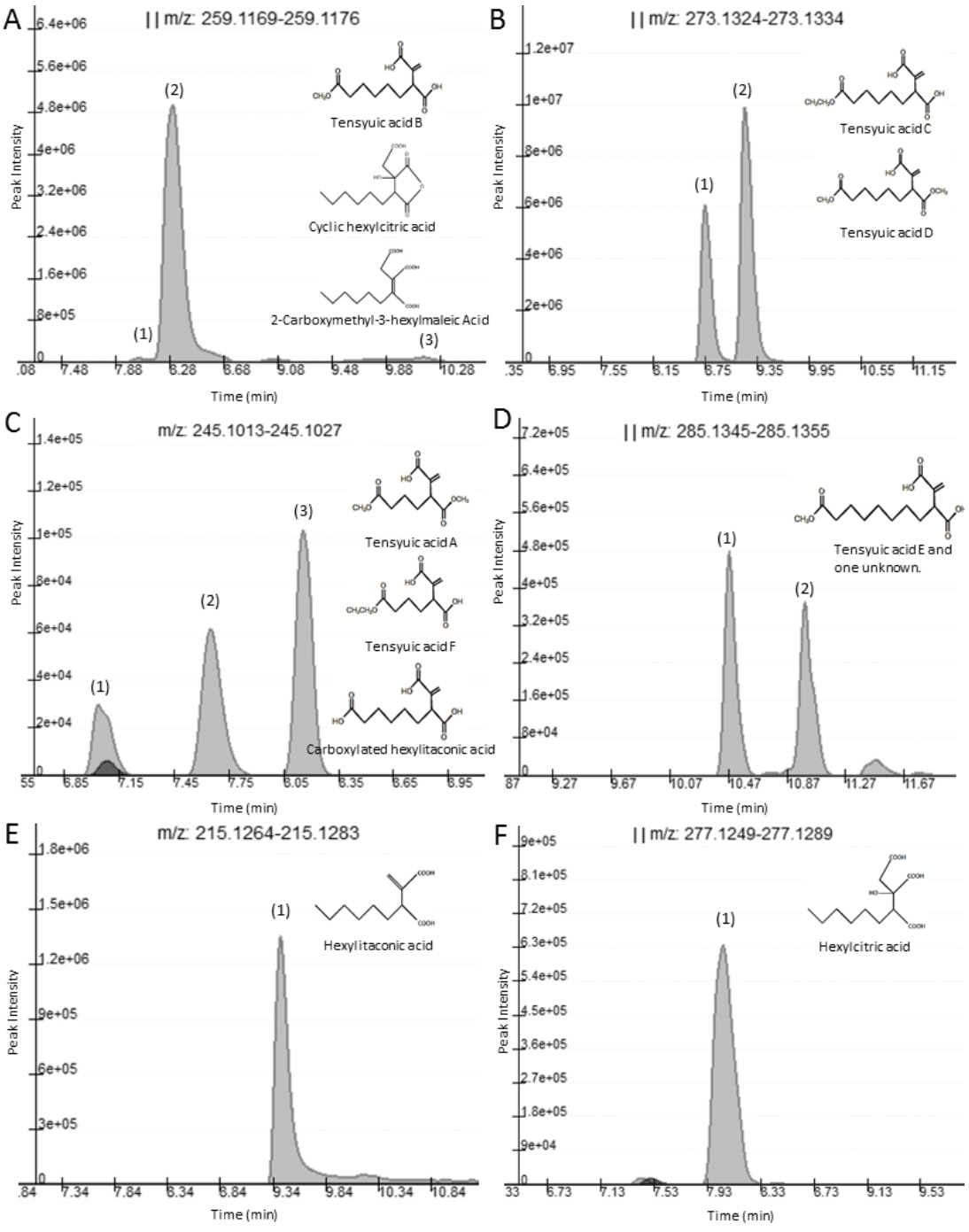
Figure 18. Structures elucidated from *chaR*^{OE} media extracts.

Table 9. Chemical shifts and coupling constants for all elucidated structures isolated from *chaR*^{OE} media extracts. Compound numbers correspond to structures in Figure 18.

No.	Chemical shifts (ppm) and coupling constants (Hz)									
	2-carboxymethyl-3-hexylmaleic acid		Hexylitaconic acid		Hexylcitraconic acid		Hydroxylated hexylitaconic acid		Carbonylated hexylitaconic acid	
	^{13}C	^{1}H	^{13}C	^{1}H	^{13}C	^{1}H	^{13}C	^{1}H	^{13}C	^{1}H
1	165.093		179.571		166.253		175.709		178.495	
2	147.95		46.288	3.485 (t) J_{2-5}: 7.0	144.755		46.561	3.451 (t) J_{2-5}: 7.0	46.506	
3	135.596		136.909		140.415		139.428		137.075	3.446 (t); J2-5: 7.5
4	165.07		171.649		165.859		168.105		170.928	
5	24.782	2.491 (t); J_{5-6}: 8.0	30.565	5a: 1.904 (m); 5b: 1.706 (m)	24.426	2.451 (dt) J_{6-7}: 7.0; J_{6-8}: 0.5	30.769	5a: 1.852 (m); 5b: 1.684 (m)	30.009	5a: 1.932 (dd); J5-2: 7.5; J5-6: 13.5 5b: 1.716 (dd); J5-2: 7.5; J5-6: 13.5
6	27.309	1.596 (m)	27.279	1.368 ÷ 1.245 (m)	27.537	1.574 (m)	27.241	1.323 ÷ 1.369 (m)	26.799	1.309 (m)
7	28.971	1.283 ÷ 1.303 (m)	28.888	1.368 ÷ 1.245 (m)	29.077	1.304 (m)	28.857	1.323 ÷ 1.369 (m)	28.922	1.342 (m)
8	31.149	1.283 ÷ 1.303 (m)	31.521	1.368 ÷ 1.245 (m)	31.331	1.304 (m)	25.328	1.323 ÷ 1.369 (m)	23.263	1.591 (m)
9	22.24	1.283 ÷ 1.303 (m)	22.529	1.368 ÷ 1.245 (m)	22.438	1.304 (m)	32.113	1.518 (m)	43.274	2.442 (t); J9-8: 7.5
10	13.718	0.882 (t); J_{10-9}: 7.0	13.992	0.874 (t) J_{10-9}: 7.0	13.969	0.889 (t) J_{11-10}: 7.0	61.556	3.532 (t) J10-9: 7.0	209.493	0.882 (t); J10-9: 7.0
11	29.1	3.566 (s)	130.231	11a: 6.550 (s);	9.477	2.069 (s)	125.625	11a: 6.329 (s); 11b: 5.757 (s)	129.851	11a: 6.527 (d); J11a-11b: 3.0 11b: 5.866 (d); J11b-11a: 3.0
12	172.56			11b: 5.910 (s)						

3.2.4 Overexpression of transcription factor *chaR*: The tensyuic acids

In both positive and negative MS modes, the *chaR*^{OE} strain exhibited peaks with masses corresponding to those of the four-hydrocarbon tail (C4) tensyuates A and F, the six-hydrocarbon tail (C6) tensyuates B-D, and the eight-hydrocarbon (C8) tensyuate E. Moreover, the order of elution (F, A, B, C, D and E) of these acids, acknowledging the ambiguity of equal massed tensyuates A/F and C/D, appears consistent with that reported by Klitgaard, A., et al. The Klitgaard, A., et al study, like in our study, used a C18 column and a similar solvent gradient to ours [142]. To help visualize these peaks in isolation between *chaR*^{OE} and the control strain, we used extracted ion chromatograms (EIC) (Figure 19A-D). One item of note was the absence of detectable C4 and C8 versions of hexylitaconic acid and 2-carboxymethyl-3-hexylmaleic acid above peak intensity threshold ($5x10^4$) indicated either a potential preference for the production of the C6 tensyuate variates in *A. niger* NRRL3. All EIC's for putative compounds identified by mass spectrometry are shown in Figure 19. The MS profile of the intracellular fraction showed a similar peak complement to that of the extracellular media. Intracellular fractions did not have any unique identified peaks and were missing peaks with m/z ratios of 245.1017 (tensyuic acids A/F), 197.117 (hexylcitratconic acid) and 257.1017 (veridicatic acid) that were present in external fractions. Peaks which were common to both fractions were ~2 orders of magnitude lower in the intracellular fractions compared to external fractions.


A
|| m/z: 259.1169-259.1176
Tensyuic acid B
Cyclic hexylcitric acid
2-Carboxymethyl-3-hexylmaleic Acid
(1)
(2)
(3)
Peak Intensity
Time (min)
B
|| m/z: 273.1324-273.1334
Tensyuic acid C
Tensyuic acid D
(1)
(2)
Peak Intensity
Time (min)
C
m/z: 245.1013-245.1027
Tensyuic acid A
Tensyuic acid F
Carboxylated hexylitaconic acid
(1)
(2)
(3)
Peak Intensity
Time (min)
D
|| m/z: 285.1345-285.1355
Tensyuic acid E and one unknown.
(1)
(2)
Peak Intensity
Time (min)
E
m/z: 215.1264-215.1283
Hexylitaconic acid
(1)
Peak Intensity
Time (min)
F
|| m/z: 277.1249-277.1289
Hexylcitric acid
(1)
Peak Intensity
Time (min)

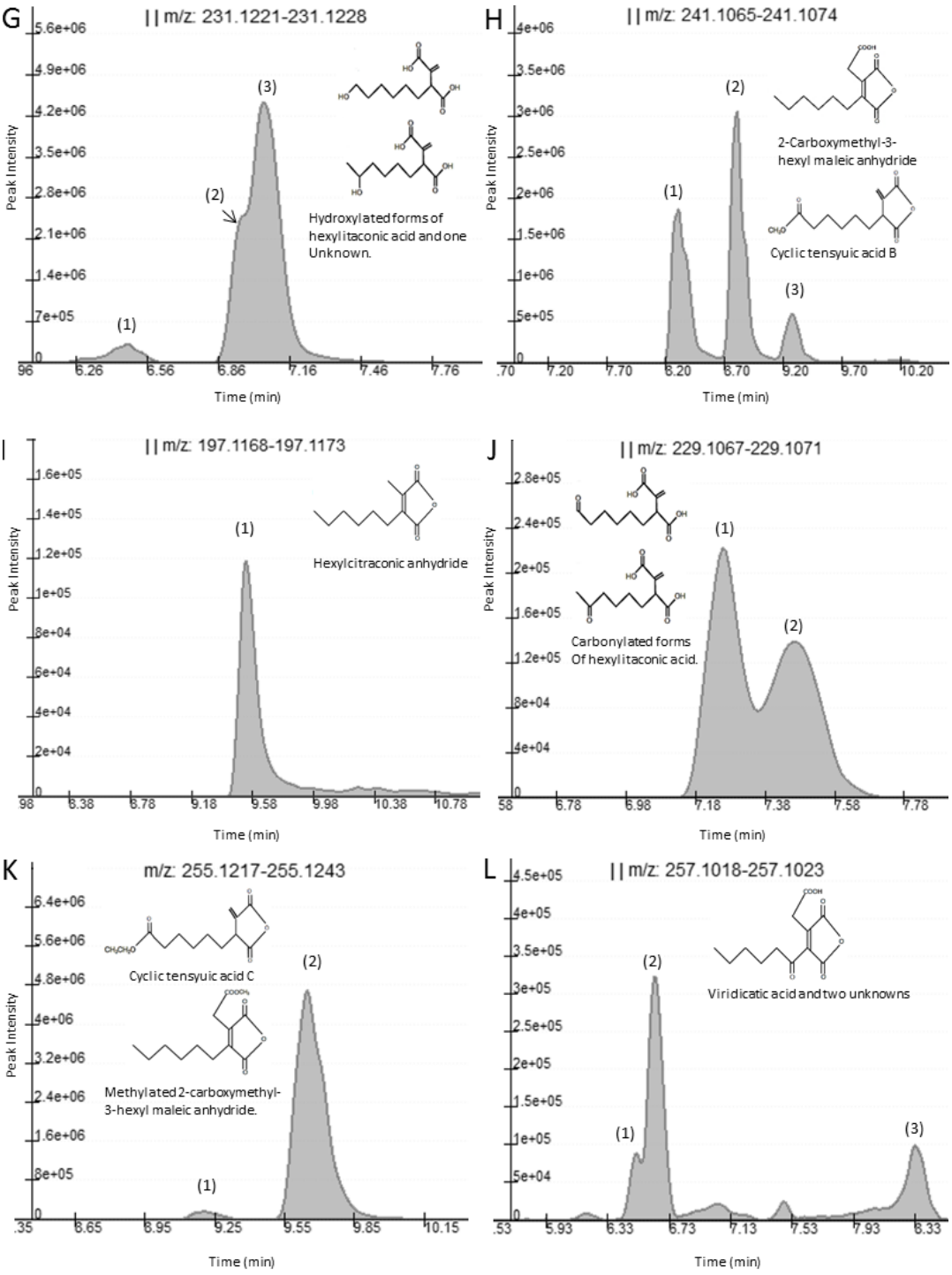


Figure 19. Overlaid EICs of control (black) and *chaR*^{OE} (grey) strains. All EICs are shown in positive mode excluding panel D in negative mode. Numbers above peaks represent peaks above detection threshold; peaks below detection threshold are not numbered apart from panel C where peak 1 appeared above threshold in negative mode only.

3.3 Alkyl citric acid pathway reconstruction in *A. niger*

Combining MS, NMR and transcriptomic data, we reconstructed a biosynthetic pathway that included the tensyuic acids. To fill in the gaps in the biosynthetic sequence for the production of 2-carboxymethyl-3-hexylmaleic acid and the tensyuic acids, I tried to anticipate the existence of intervening compounds. Adding the chemical formulas of predicted compounds to our database, I attempted to identify previously unidentifiable peaks in our MS spectrum data to gain supporting evidence for the compounds presence in growth media. Starting with 2-carboxymethyl-3-hexylmaleic acid, we anticipated the existence of a 2-hexylcitric acid precursor based on the up-regulation of *cha6,* the gene predicted to encode 2-methylcitrate dehydratase (Table 7). A lactone ring anhydride form of 2-hexylcitric acid was also anticipated since a ring forms of other alkyl citrates had also been previously identified [92]. The presence of a ring opening dienelactone hydrolase (the up-regulated *cha2*, Table 7) also supported this prediction. Surveying our MS data with the chemical formulas for these two compounds, peaks with m/z ratios corresponding to 2-hexylcitric acid and its anhydride were identified in *chaR*^{OE} spectra (Figure 19A, F).

For an explanation of 2-hexylcitric acid/anhydride biosynthesis, we attempted to distinguish whether the FAS is linking a completed fatty acid to a citric acid in the last iteration or if the FAS uses a citric-acid derivative as a starter unit. Examining scenarios where citric acid is attached at the end or the beginning of the FAS chain elongation process failed to infer 2-hexylcitric acid production due to the necessary loss of a carboxylic acid group in the form of CO_2 from the citric acid moiety of 2-hexylcitrate in the condensation reaction [83, 143]. The loss of a carboxylic acid group from the fatty acid would result in a molecule with odd numbered hydrocarbon tail which was not detected in our MS spectrum data. Attention shifted to the citrate synthase (Cha5) enzyme as a possible medium to carry out a condensation reaction. Analogous to the mechanism in citrate, 2-methylcitrate and 2-decylcitrate synthase enzymes [144, 145], I proposed that the clustered citrate synthase acts as the linking enzyme for oxaloacetate and fatty acid-CoA. Supporting this idea, our expression data indicate a significant up-regulation of two citrate lyase subunit genes (NRRL3_10102, NRRL3_10103) whose enzyme product breaks down cytosolic citric acid to oxaloacetate and acetyl-CoA (Table 5) [146]. During the experimental work for this thesis, a similar mechanism using a citrate synthase and 2-

methylcitrate dehydratase enzyme was also proposed for the similarly structured hept/nonadride alkyl citrates by two other groups in the fungi *Talaromyces stipitatus* and *Byssochlamys fulva* [147, 148].

For the biosynthesis of hexylitaconic acid, I proposed a decarboxylation of 2-carboxymethyl-3-hexylmaleic acid; analogous to the itaconic acid pathway [141]. Reviewing our transcriptomic data (Table 7), I identified a non-clustered and significantly up-regulated cis-aconitate decarboxylase gene (NRRL3_00504) which may be carrying out the decarboxylation reaction however, the clustered gene of unknown function (NRRL3_11757) cannot be ruled out as a candidate for this reaction. Alternatively, the recent work with the similarly structured hept/nonadrides suggested that the decarboxylation reaction may be occurring non-enzymatically via a "facile decarboxylation" [149].

Focusing on the ring and open forms of the alkyl citrates, I surmised that the *cha5* citrate synthase produces the ring anhydride forms of the alkyl citrates instead of the open forms for two reasons. Firstly, the Cha2 enzyme facilitates the conversion of anhydride to acid forms and there does not appear to be any upregulated genes in our expression data which could produce enzymes to facilitate the reverse reaction. Secondly, spontaneous condensation reactions of carboxylic acid converting dicarboxylate molecules to closed anhydride forms do not occur to the best of our knowledge [150]. It is currently unclear if the subsequent reactions converting 2-hexylcitric acid/anhydride to the tensyuates take place in the open, closed or both forms. Figure 20A shows a reconstructed pathway for 2-carboxymethyl-3-hexylmaleic acid and hexylitaconic acid.

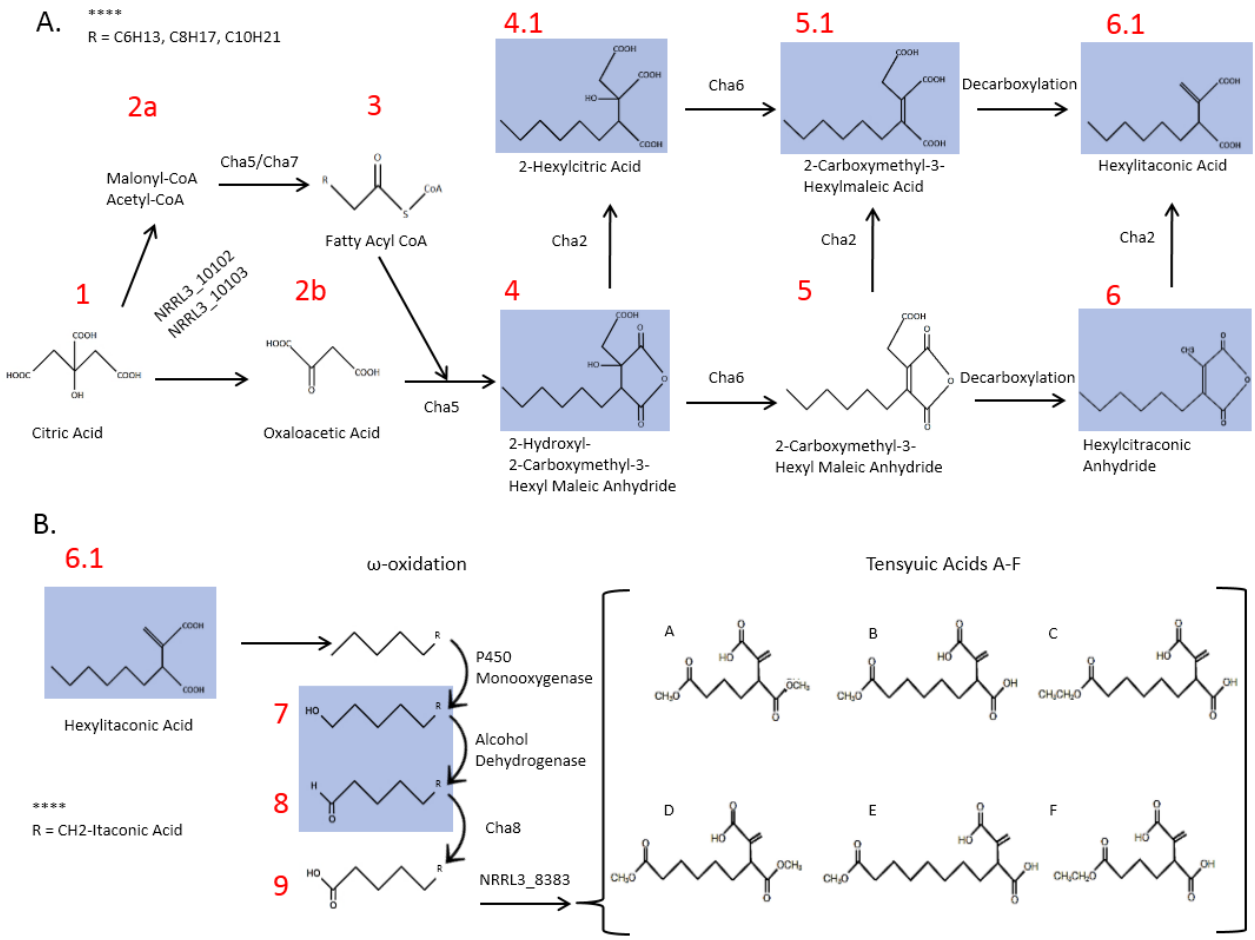


Figure 20. A proposed biosynthetic pathway for hexylitaconic acid (A) and its conversion into the tensyuic acids (B). Shaded boxes represent compounds identified by both MS and NMR. All non-shaded compounds, apart from compounds 2a, 2b and 3, (inferred by enzymatic actions of FAS and citrate lyase enzymes) were identified by MS only.

The lack of enzymes in the *cha* cluster to carry out tensyuate biosynthesis suggested that non-clustered enzymes might be involved; like we observed with the up-regulated and non-clustered citrate lyase genes. The transcriptomic data revealed a single significantly up-regulated o-methyltransferase (NRRL3_08383, only expressed in *chaR*^{OE} strain) which may be carrying out the o-methylations/o-ethylations (o-(m)ethylations) generating the tensyuates. As for the carboxylate tail, the hydroxylated and carbonylated forms of hexylitaconic acid found in our NMR and MS analysis as well as the clustered aldehyde dehydrogenase (*cha8*) provided potential clues. I surmised that a ω-oxidation reaction might be taking place on the terminal end of the hexylitaconic acid hydrocarbon tail. A key piece of evidence was clustered aldehyde

dehydrogenase which facilitates the terminal enzymatic step in the ω-oxidation reaction. Additionally, the ω-oxidation reaction includes a P450 monooxygenase generating a hydroxylated hydrocarbon tail [148] and an alcohol dehydrogenase generating a carbonylated hydrocarbon tail [151]. This pathway has also been previously characterized in filamentous fungi with other fatty acid compounds [152]. Transcriptomic data showed two monooxygenases (NRRL3_03003, NRRL3_03750) that were significantly upregulated in the *chaR*^{OE} strain (Table 7). For the dehydrogenation portion of the ω-oxidation reaction, twelve zinc-type alcohol and short chain dehydrogenase candidates are significantly upregulated one or more of which may be contributing to the ω-oxidation reaction (Table 7). Figure 20B shows a proposed pathway converting hexylitaconic acid to the tensyuates. Lastly, we detected a number of peaks with m/z ratios corresponding to viridicatic acid; a compound nearly identical to 2-carboxymethyl-3-hexylmaleic anhydride possessing an additional keto group near the aconitate head. The other peaks may indicate alternate locations of the keto group including at the terminal end of the hydrocarbon tail potentially as part of a ω-oxidation reaction of 2-carboxymethyl-3-hexylmaleic anhydride.

3.4 Supporting the reconstructed pathway by gene deletion: Using the reconstructed alkyl citric acid pathway to predict the metabolic outcomes from gene deletions

To test the predictive power of our reconstructed pathway we designed three deletion cassettes for three genes in our pathway. We selected first the clustered *cha6* gene for deletion to lend support for the existence of our predicted 2-hexylcitric acid/anhydride and potentially obtain a structural solution by NMR. Attributing the inability to detect 2-hexylcitric acid/anhydride to a lack of sufficient material, we anticipated that a *cha6* deletion strain (Δ*cha6*:*chaR*^{OE}) would achieve 2-hexylcitric acid/anhydride accumulation and allow for a structural determination of these compounds by NMR. I also anticipated the elimination of compounds downstream of 2-hexylcitric acid/anhydride in the pathway.

The Δ*cha6*::*chaR*^{OE} strain yielded an increased peak area corresponding to the mass of 2-hexylcitric acid with a $\log_2$ fold change of 2.982 ± 0.15 (p-value < 0.01) and an increased peak area corresponding to the mass of 2-hexylcitric anhydride with a $\log_2$ fold change of 3.792 ± 0.26 (p-value < 0.01). All downstream compounds including the tensyuate m/z peaks were not

detected in MS data (Figure 21A-D). Analyzing media extracts of our Δ*cha6*::*chaR*^{OE} by NMR we were able to verify the structures of 2-hexylcitric acid and 2-hexylcitric anhydride (Table 10, Figure 22).

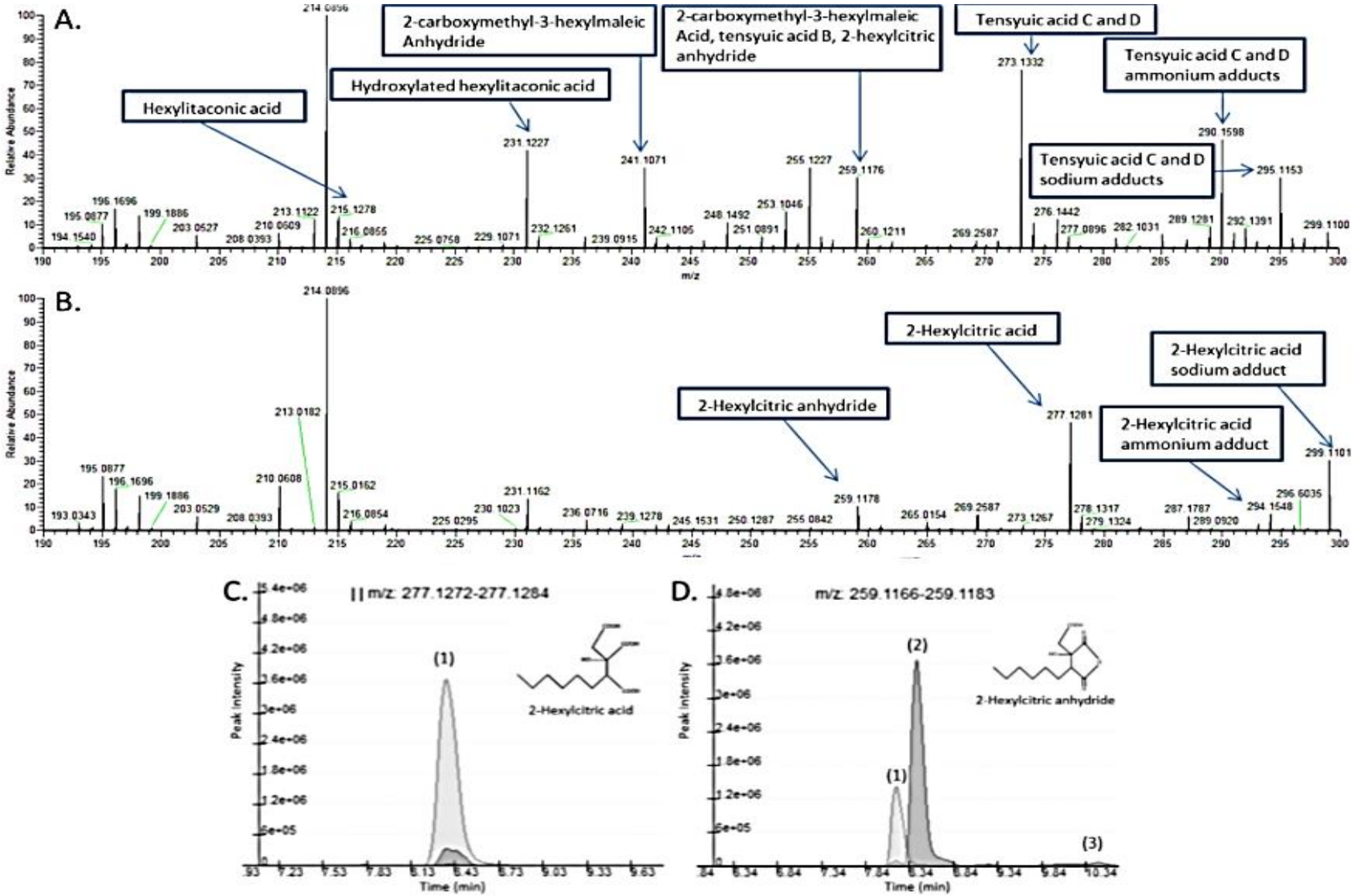
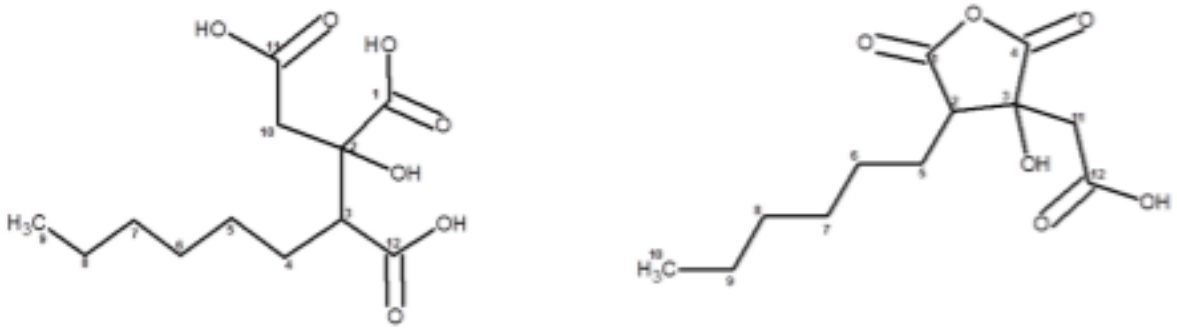


Figure 21. Mass spectra of *chaR*^{OE} strain (A). Mass spectra of Δ*cha6:chaR*^{OE} (B). Overlaid EICs from the *chaR*^{OE} (dark gray), Δ*cha6:chaR*^{OE} (light gray) and control strains (black but too low for visibility) for 2-hexylcitric acid in positive mode (C), 2-hexylcitric anhydride in positive mode (D). The 2-hexylcitric anhydride compound appears to correspond to peak 1 in panel D.

Table 10. Chemical shifts and coupling constants for 2-hexylcitric acid and 2-hexylcitric anhydride structures isolated from Δ*cha6::chaR*^{OE} media extracts.

No.	Chemical shifts (ppm), coupling constants (Hz), and COSY correlation					
	2-hexylcitric acid			2-hexylcitric anhydride		
	^{13}C	^{1}H	COSY	^{13}C	^{1}H	COSY
1	DMSO *	1a: 2.898 (d); 1b: 2.504 (d); J_{1a-1b}: 16.0	H1a-H1b			
2	79.510			173.932		
3	DMSO *	2.713 (dt) J_{3-4a}: 14 ; J_{3-4b}: 1.0	H3-H4a-H4b	75.458		
4	31.460	4a: 2.377 (m); 4b: 1.651 (m)	H4a-H4b-H5	54.081	2.4025 (m)	H4-H6a-H6b
5	27.112	1.444 (m)	H5-H4-H6	175.215		
6	29.032	1.226 (m)	H6-H5-H7	26.923	6a: 1.651 (m), 6b: 1.393 (m)	H6a-H6b-H7a-H7b
7	31.582	1.226 (m)	H7-H6-H8	27.750	7a: 1.226 (m), 7b: 1.137 (m)	H7a-H7b-H6a-H6b-H8
8	21.208	1.118 (m)	H8-H9-H7	28.941	1.226 (m)	H8-H7a-H7b-H9
9	14.531	0.842 (t) J_{9-8}: 7.0	H9-H8	31.460	1.226 (m)	H9-H8-H10
10	174.183			22.423	1.226 (m)	H10-H11-H9
11	174.334			14.349	0.847 (dt) J_{11-10}: 7.0, J_{11-9}: 1.5	H11-H10-H9
12	178.022			41.227	12a: 2.898 (d); 12b: 2.504 (d); $J_{12a-12b}$: 16.0	H12a-H12b
13				171.974		

*indicates signal overlap with DMSO.



2-Hexylcitric acid

2-Hexylcitric anhydride

Figure 22. Structures elucidated from Δ*cha6::chaR*^{OE} media extracts.

Two additional cassettes were constructed to test the ω-oxidation and o-(m)ethylation aspects of the reconstructed pathway. I selected the NRRL3_03750 as a deletion target to test ω-oxidation since the $\log_2$ fold change in expression for NRRL3_03750 was higher than for NRRL3_03003 (4.95 ± 0.15 vs 2.78 ± 0.02). The o-methyltransferase NRRL3_08383 was selected to test o-(m)ethylation in the pathway as the only candidate for that reaction. Based on our pathway reconstruction, we predicted that a ΔNRRL3_03750::*chaR*^{OE} strain should terminate production at step 6 (Figure 20B) while a ΔNRRL3_08383::*chaR*^{OE} strain would terminate production before the tensyuic acids at step 8 and potentially generate enough material for NMR structural resolution of the carboxylated hexylitaconic acid at that step.

The MS profile of the ΔNRRL3_03750::*chaR*^{OE} strain showed no apparent difference in identified peaks when compared to the MS profile of *chaR*^{OE}. Based on these data NRRL3_03750 does not appear to be involved in alkyl citric acid production. A deletion cassette was also constructed for the o-methyltransferase NRRL3_08383 however the ΔNRRL3_08383::*chaR*^{OE} strain could not be found despite extensive mutant screening.

3.5 Citric acid levels in control, *chaR*^{OE} and Δ*cha6:chaR*^{OE}

To support our premise that citric acid is the primary molecule used for alkyl citric acid production, we measured the relative fold change of citric acid in control, *chaR*^{OE} and Δ*cha6::chaR*^{OE} strains. We predicted that both the *chaR*^{OE} and Δ*cha6::chaR*^{OE} strains would

require a greater pool of citric acid than the control strain in order to sustain both alkyl citric acid production and primary metabolism from the citric acid cycle. We also noted that the *chaR* transcription factor contains a mitochondrial sequence tag indicating that it may also be involved in the regulation of the citric acid cycle. The fold change in citric acid levels was determined by peak area from MS data. We observed a significant increase of citric acid in the $chaR^{OE}$ ($\log_2$ fold change = 5.36 ± 0.81, p-value < 0.01) and Δ*cha6::*$chaR^{OE}$ ($\log_2$ fold change = 3.96 ± 0.77, p-value < 0.01) strains compared to the control strain. Furthermore, the level of citric acid in the $chaR^{OE}$ strain was significantly higher than in the Δ*cha6:*$chaR^{OE}$ strain ($\log_2$ fold change = 1.41 ± 0.39, p-value < 0.01).

Discussion

4.1 Secondary metabolism genes and gene clusters in *A. niger* NRRL3 genome

4.1.1 Backbone enzymes in the genomes of *A. nidulans*, *A. oryzae*, *A. fumigatus* and *A. niger*.

Differences in the distribution of SM enzyme types found in *A. nidulans*, *A. oryzae*, *A. fumigatus* and *A. niger* indicate that in addition to evolutionary relatedness, determination of the complement of SM genes in *Aspergillus* species may be subject to both horizontal gene transfer and gene cluster loss. We observed in the case of the four *Aspergillus* species under study that the distribution of SM enzyme types in *A. nidulans* appears to be more similar to that of *A. niger* and *A. oryzae* than that of the more closely related *A. fumigatus* [152]. The *A. fumigatus* genome also appears to have comparatively very few SM backbone enzymes compared to the other *Aspergilli* suggesting substantial SM gene cluster loss during its evolution. While a phylogenetic and sequence analysis of the backbone enzymes in the studied *Aspergilli* was not performed to see which SM enzymes are shared between them, the number and category distribution differences of these enzymes might be explained by both horizontal gene transfer and gene/cluster loss. Horizontal gene transfer of SM clusters has been previously inferred in filamentous fungi in the phylogenetic distributions of the sterigmatocystin and fumonisin clusters [152, 153].

An additional line of evidence for the dynamic nature of SM clusters comes from our *A. niger* species comparison. The results of this analysis showed, as in previous studies, that strains of the same species can lose or gain not just backbone enzymes but entire SM clusters (Figure 16) [75]. In some cases a remnant backbone "gene" potentially indicates where an SM cluster had left the genome.

4.1.2 Remnant cluster analysis

The BLASTP method using the query sequences of contiguous SM domains to search for backbone enzymes in *A. niger* NRRL3, helped locate remnant backbone genes (in addition to the ochratoxin PKS remnant) in both the *A. niger* NRRL3 and CBS 513.88 genomes. One aspect of this search revealed that it is mainly the highly reducing PKS backbone class that tend to leave behind a gene remnant. To our knowledge, the reason and mechanism behind this observation is

currently unknown but a potential clue may lie in the evolutionary relatedness of highly reducing PKS enzymes with fungal FAS backbone genes. Both FAS and PKS enzymes are currently thought to have a common point of origin given the similarities in their domain structure and metabolic products [154]. The tendency of highly reducing PKS enzymes to leave behind a portion of their 3' end sequence may be an indication of their evolutionary origins from a progenitor it shared in common with fungal FASs. A typical fungal FAS exists as a heterodimer wherein the alpha subunit contains an ACP domain followed by KR and the beta subunit containing AT and ER domains [153]. The tendency for PKS-like remnants to retain the ACP, and in some cases the KR domain, but not AT and rarely ER (An01g02030) domains may point to an FAS-like enzyme subunit fusion event in the past that later evolved into the highly reducing PKSs. Further analysis and testing of these remnants from within other genomes are required to obtain a more extensive list of PKS-like enzymes to support our limited findings.

Using protein sequences with only single SM domains as a subsequent BLASTP remnant search, no additional remnant genes were found. The vast majority (89 of 114) of single domain proteins observed in the NRRL3 genome were the PKS ER and KR domains. Moreover, only one instance of a PKS-KS and PKS-AT domain were observed and no instances of genes containing a single PKS-DH could be found. The significance of this observation is currently unclear however I speculate that these single domains are modified over time from remnants since nearly all of the single domain proteins sequences I could find appear in the 3' end of highly or partially reducing PKSs like the remnants. These proteins may have been adapted for other novel functions or potentially as tailoring enzymes in other SM clusters. A more extensive search through other fungal genomes may provide support for this hypothesis.

Lastly, the PKS-like remnants appeared in two contexts within the genome. Either only the backbone enzyme was missing or surrounding clustered genes and the backbone gene were missing. It is unclear why these two types of cluster loss occur. A detailed study search for common sequence markers from within intact clusters and surrounding a homologous remnant may provide insight as to a possible mechanism of SM cluster loss by revealing the proteins which bind to those sequences.

4.1.4 The azanigerone and unknown FAS split cluster

The position of the azanigerone cluster within the genome is one of the puzzling aspects of our cluster analysis. To our knowledge, there is no observed instance of SM clusters appearing within other separate SM clusters. It is currently unclear if the azanigerone cluster inserted itself within the FAS cluster or if the FAS cluster has split and inserted at either end of the azanigerone cluster. A phylogenetic analysis of both clusters in related species may help sort out which scenario is correct. If most related species possess the azanigerone cluster alone it is more likely that the FAS cluster inserted around it. Conversely if most related species have the FAS cluster alone it is likely that the aznigerone cluster inserted into the FAS cluster. This unique organization may also present an insight into the horizontal gene transfer of SM clusters. By analyzing the sequences of these split clusters, DNA sequence markers could be revealed which may serve to identify mechanisms for SM cluster insertions (as opposed to the deletions outlined in the preceding section) into fungal genomes.

4.1.5 Backbone definition in *A. niger* NRRL3 genome

A comparison of the clusters defined in *A. niger* NRRL3 to CBS 513.88 revealed both strains share a nearly identical set of SM clusters. However, like in previous *A. niger* strain comparisons [75], we also observed some differences in the composition of some clusters. Cluster differences between *A. niger* NRRL3 and CBS 513.88, as well as between other *Aspergilli*, appeared to take four forms: 1) completely missing clusters, 2) rearranged or dispersed clusters, 3) clusters with absent tailoring enzymes, and 4) truncated/remnant backbone genes. Differences corresponding to tailoring gene complement within clusters or cluster rearrangements appear to have only minor effects on SM production in some cases. The fumonisin cluster of *A. niger* has only 11 of the 17 genes reported in *Fusarium* species but can still produce fumonisins B2 and B4 [153]. Additionally, the scattered pigmentation cluster in *A. niger* still generates DHN melanin and its pigment derivatives despite its extensive rearrangement in the genome [103]. In cases where a backbone enzyme is missing, metabolomics changes can be much more profound. For example, the inability to produce ochratoxin by strain NRRL3, as compared to strain CBS 513.88 [113] can be explained by partial loss of the ochratoxin PKS gene since the core molecule can no longer be made.

4.2 Overexpression of defined clusters

4.2.1 Transcription factor overexpression induces the production of SMs from their parent clusters in only a few cases

By overexpressing the regulators of defined clusters, we sought to determine their metabolic products. In the majority of cases, MS data from strains overexpressing clustered transcription factors (28 of 35) were similar to the spectra from the non-expressing controls indicating a lack of expression from cryptic clusters. The reason for the lack of cluster activation by clustered transcription factor overexpression is unclear. I submit however, four potential explanations that may serve as considerations in similar experiments in the future. First, given the random integration of our overexpression vectors, the observed difficulties in cluster expression might be linked to vector insertion into transcriptionally silent or poorly expressing areas of the genome. Similarly, insertion in areas of high expression may cause a negative feedback and shut down production of the transcription factor. A targeted insertion of transcription factor overexpression vectors into moderately transcriptionally active areas may solve this problem. Second, some mechanism of repression may be involved in the regulation of SM clusters. The relatively low expression level of *chaR* despite being under the control of a strong promoter could indicate that *A. niger* may have countermeasures in place to reduce SM transcription factor transcript/protein levels or to inhibit their function. Deletion of these inhibitory mechanisms, once identified, may facilitate SM biosynthesis from target clusters. Third, clustered transcription factors may only regulate a small number of SM clusters and the majority may be regulated by transacting transcription factors from elsewhere in the genome or by global regulators like the LaeA which regulates 20-40% of SM genes in *A. fumigatus* [109]. Global repressors of secondary metabolism may also be considered. Lastly, epigenetic factors can also influence the expression of SM clusters [110]. Reducing epigenetic gene silencing mechanisms may then induce expression either on its own or by subsequent expression of clustered transcription factors.

4.2.2 Transcription factor induced SM production reveals the putative genetic underpinnings of orphan compounds

From the transcription factors that appeared to induce SM production, we detected the overproduction of three identifiable orphan compounds. These were pyrophen, the alkyl citric acids and the malformins. The gene clusters associated with pyrophen and the tensyuic acids appear to have enzymatic functions that coincide with the chemical structures for these compounds. In the case of the putative pyrophen cluster NRRL3_00410 (transcription factor NRRL3_00406), a tyrosinase/chatechol oxidase and an NRPS-like gene provide ready explanations for the incorporation of a benzyl group linked to a nitrogen atom from an aromatic amino acid like tyrosine or phenylalanine (Figure 9, Table 4). Looking at the alkyl citric acids, the FAS and citrate producing/modifying enzymes in the NRRL3_11765 cluster appeared to explain the fatty acid and citrate derived components of the tensyuic acid compounds [93]. In the case of the malformins however, the biosynthetic mechanism could not be worked out based on the cluster in the vicinity of the transcription factor.

Overexpression of NRRL3_07881 transcription factor resulted in the over-production of malformins. However, the SM gene cluster encompassing the NRRL3_07881 transcription factor does not show any corresponding backbone or tailoring enzymes explaining its biosynthesis. The malformins are cyclic pentapeptides containing a disulfide bridge between two cysteine residues (Figure 9). The main backbone enzyme in the NRRL3_07881 cluster is a partially reducing PKS (NRRL3_07884) which does not generate peptides according to current understanding [50, 67, 71]. An NRPS-like enzyme (NRRL3_07881) is also present but it contains only a single A domain. Since the current understanding of NRPS peptide assembly proposes that each A domain can recognize only one amino or hydroxyl acid and because malformins comprise four different amino acids (L-isoleucine, two D-cysteines, L-valine, and D-leucine), the minimum number of A domains for a malformin producing NPRS should be four [155]. Moreover, given that both D-amino acids are incorporated along with L-amino acids, ribosomal production of the malformins is also unlikely given that protein biosynthesis only incorporates L-amino acids and does not cause peptide cyclization. Considering these constraints, NRRL3_07881 may be regulating malformin production from a distantly located cluster. A review of the annotated clusters revealed that, an NRPS cluster containing the NRPS backbone NRRL3_08969 is a likely

candidate. The NRPS backbone contains four A domains as predicted as well as a disulfide modifying enzyme (NRRL3_08968) which could account for the disulfide bridge between malformin's two cysteine residues (Figure 9B) [156]. Proteomic analysis (data not shown) of our malformin producing strain was inconclusive as no backbone enzymes were found to be up regulated. A transcriptomic profile may provide a more definitive answer.

4.3 Alkyl citric acid pathway reconstruction in *A. niger*

4.3.1 Low production of some metabolites may preclude their structural elucidation by NMR.

The pathway reconstructed by NMR, MS and transcriptome data provided insights into the biosynthesis of the alkyl citric acids. However, some of the proposed precursors and terminal products could not be identified by NMR. We attributed our inability to obtain structural solutions for these compounds to a lack of sufficient material. Looking at 2-carboxymethyl-3-hexylmaleic anhydride, the lack of sufficient material may be due in part to the ring opening action of the *cha2* gene ($\log_2$ fold change = 6.11 ± 0.90) skewing the state of this compound to an open acid form. As for the tensyuic acids and the tail carboxylated hexylitaconic acid, the lack of sufficient material seemed less obvious. The *cha* cluster lacked all enzymes, save *cha8* ($\log_2$ fold change = 2.77 ± 0.36), to carboxylate the hexylitaconic acid tail by ω-oxidation. However, the low expression levels of the gene *cha8*, the gene of unknown function (which may also be involved in ω-oxidation) ($\log_2$ fold change = 3.41 ± 0.23) and the non-clustered cis aconitate decarboxylase NRRL3_00504 ($\log_2$ fold change = 2.70 ± 0.02) may point to a low frequency of ω-oxidation and subsequent decarboxylation due low enzyme production. Based on our pathway, this could then lead to a low concentration of tensyuic acids.

It is also worth pointing out that some of the compounds which could not be identified by NMR had MS peak areas comparable to or higher than some of those which could be elucidated by NMR (Figure 19). For instance, peak areas corresponding to tensyuic acids C and D (Figure 19B) that could not be identified by NMR were ~10 fold higher than that of hexylitaconic acid (Figure 19E) which could be identified by NMR. This discrepancy may be due to ionization differences between these compounds; those which ionize better may produce higher peaks despite having low concentrations since more of its ions hit the MS detector.

4.3.2 Strategies for the detection of the tensyuic acids

Currently the only support this study has been able to show for the production of the tensyuates by *chaR* overexpression is the detection of peaks with m/z ratios corresponding to the tensyuic acids in MS data and the order of their elution on a C18 column. If a lack of sufficient material is responsible for our inability to detect the tensyuates by NMR, increasing tensyuate concentration may help. To achieve this, we propose two ways of increasing tensyuate concentration. The first method involves the overexpression of the decarboxylating enzyme(s) generating increased hexylitaconic acid or the enzymes putatively involved in ω-oxidation (like *cha8*). A second method is to simply grow greater volumes of stationary culture to isolate enough material.

4.3.3 Citric acid feeds the alkyl citric acid pathway

Based on the upregulated citrate lyase subunits which break down citric acid [146] and the action of the Cha5 citrate synthase, citric acid appears to be the primary building block feeding the alkyl citric acid biosynthetic pathway. Assays to quantify citric acid by MS analysis also showed that transcription factor induced expression of this pathway appears to have a corresponding increase in the concentration of extracellular citric acid. In the case of the Δ*cha6:chaR*^{OE}, we observed that while the relative concentration of citric acid was significantly higher than the control strain, it was significantly lower than the *chaR*^{OE} strain. The reason for this difference is unclear however, I speculate that the significant increase of 2-hexylcitric acid/anhydride may be causing a negative feedback loop through some as yet unknown mechanism which is either reducing citric acid production, reducing the transport of citric acid to the cytosol or shunting the breakdown products (oxaloacetic acid and acetyl-CoA) of citric acid into other pathways.

A key feature in this early part of the alkyl citric acid pathway appears to be a highly upregulated mitochondrial citrate/oxoglutarate carrier protein (NRRL3_00546, $\log_2$ fold change = 2.58 ± 0.23) (Table 7). In *Saccharomyces cerevisiae*, the orthologous mitochondrial transporter functions as a citric acid shuttle (by oxoglutarate exchange) from the mitochondrion to the cytosol in order to increase the redox state of cytosolic NADPH (Figure 23) [157]. The specific role of this antiporter in the reconstructed alkyl citric acid pathway is unknown as the enzymes

converting citrate to isocitrate (Aco1p) and isocitate to oxoglutarate (Idp2p) do not appear to have a corresponding upregulation in the genome following *chaR* overexpression. However, I anticipate that the NRRL3_00546 up-regulation may be a response to increase the shuttling of citric acid from the mitochondria to the cytosol in order to feed the alkyl citrate pathway, generate NADPH required for fatty acid biosynthesis [158] or both. Lastly, the *chaR* transcription factor appears to contain a mitochondrial localization sequence indicating a potential role in gene regulation in the mitochondria. Green fluorescence protein fusion with *chaR* may help confirm its localization following translation.

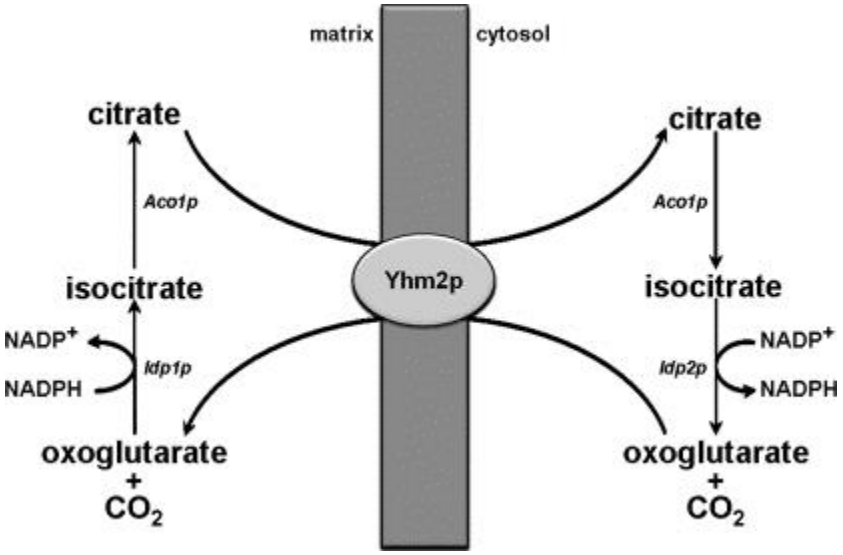


Figure 23. Schematic of mitochondrial citrate/oxoglutarate carrier protein (Yhm2p) function in *S. cerevisiae*. Taken from Castegna, A., et al. [157]

The observed significant downregulation of what appears to be an itaconic acid pathway in the *chaR* strain is also noteworthy. Since this pathway facilitates the conversion of citrate to aconitate to itaconate [141] it appears that *A. niger* may be downregulating this pathway to focus citrate flux to the alkylcitrate pathway. If this assessment is correct, deletion of the *cha* cluster may generate cytosolic citric acid and other small molecules (acetyl-CoA and oxaloacetate) with lower diversion to other pathways. A strain overproducing these small molecular building blocks could be used to make a variety of endogenous and non-native SMs; particularly fatty acid and polyketides which require acetyl-CoA and malonyl-CoA for biosynthesis. Moreover,

the upregulated acetyl-CoA carboxylase gene NRRL3_03431 (Log_2 fold change = 2.88 ± 0.57) (Table 7.) may be facilitating increased production of malonyl-CoA [159].

4.4 Testing the reconstructed pathway by gene deletion

4.4.1 The mechanism of 2-hexylcitric acid and 2-carboxymethyl-3-hexylmaleic acid biosynthesis is similar to that of the heptadride/nonadrides

Our reconstructed pathway successfully predicted the metabolic outcome of a *cha6* deletion and allowed for the structural elucidation of 2-hexylcitric acid/anhydride by NMR. During the preparation of this thesis, the citrate synthase/2-methylcitrate dehydratase reconstructed mechanism for 2-hexylcitric acid and 2-carboxymethyl-3-hexylmaleic acid biosynthesis was proposed by two other groups for the production of the hept/nonadrides in *T. stipitatus* and *B. fulva* [147, 149]. A notable difference between our alkyl citric acid pathway and the hept/nonadride pathway is in the production of the hydrocarbon tails. In the hept/nonadride pathway the tail is generated by a highly reducing PKS rather than a FAS [147]. While the full gene cluster of the hept/nonadrides does not appear to be experimentally defined in *T. stipitatus* or *B. fulva*, these clusters appear to lack dienelactone hydrolase [147, 148]. This may indicate a preference for the ringed anhydride forms (the authors only show the ringed forms) for these species while in *A. niger*, the similarly structured alkyl citric acids appear in both the acid and anhydride forms. There is currently no indication that tail carboxylation takes place for the hept/nonadrides.

4.4.2 Deletion of non-clustered genes predicted to be involved in ω-oxidation and o-(m)ethylation had no effect on alkyl citric acid production

We intended to demonstrate direct involvement of non-clustered genes in the biosynthesis of alkyl citric acid SMs by deletion of significantly upregulated monooxygenase and o-methyltransferase genes. In the case of the NRRL3_08383 o-methyltransferase gene, a deletion strain could not be obtained despite extensive screening of transformants. Additional attempts at obtaining a deletion mutant need to be carried out to support the o-(m)ethylation aspect of the reconstructed pathway.

Support for the ω-oxidation of hexylitaconic acid portion of the pathway is provided by the structural elucidations by NMR and m/z peaks in MS data of two of the three compounds in this reaction; the fatty alcohol and fatty aldehyde forms of hexylitaconic acid. Support for the hexylitaconic acid with a carboxylated tail is shown in the mass spectrum of *chaR*^{OE} media extracts (Figure 19C). The predicted accumulation of hexylitaconic acid through the deletion of the NRRL3_03750 monooxygenase failed to affect a change in biosynthesis however, the list of potential monooxygenases has not been exhausted. The significantly upregulated P450 monooxygenase NRRL3_03003 is another prospective candidate but the list of potential candidates may not be limited to only those genes which are upregulated by *chaR* upregulation. *Aspergillus niger* may be using ω-oxidation P450 monooxygenases and other enzymes which are not affected by *chaR* overexpression but which are already highly expressed.

Conclusions and future directions

Using the domain doublet BLAST search method, we have outlined a robust method in identifying backbone enzymes in fungal genomes. While additional testing in other genomes is required, this method may lend itself to automation by coding its rules into search algorithms. As for the clusters in which the backbones reside, SMURF and antiSMASH predictions appear to overlap well with experimentally verified clusters. One issue with these algorithms is that they generate clusters that are contiguous. This condition does not hold in many cases as evidenced by the many experimentally verified non-contiguous clusters. This results in cluster predictions which do not account for potential cluster discontinuity like that observed in the pyranonigrin E carlosic acid, and alkyl citric acid clusters [47, 99]. Additionally, the tendency of SMURF and antiSMASH to overestimate the size of clusters currently necessitates inclustion of manual curation in cluster definition and suggests a need for additional training sets. Ideally, the incorporation of other data sources in these algorithms, such as genome wide expression data, may aid in cluster definition by identifying differences in expression between genes in and outside of clusters. For example, an SM cluster annotated with three P450 monooxygenases may be refined to included just one if two of those genes are highly expressed while the one closest to the backbone is very poorly expressed like the rest of the cluster.

I have used the pathway I reconstructed from MS data for 2-carboxymethyl-3-hexylmaleic acid, hexylitaconic acid and their downstream products I was able to predict the metabolic outcome of a 2-methylcitrate dehydratase deletion. Based on my results, the process of pathway reconstruction I demonstrate in this study can potentially serve as a predictive aid to overproduce specific and potentially novel compounds for structural elucidation and bioactivity assays. An incipient pathway can also be useful in targeted deletions where the end goal is to increase the concentration of particular compounds.

Not all proposed compounds based on the masses observed in our MS data could be identified in NMR and therefore still cast doubt on some portions of the reconstructed pathway; namely the full ω-oxidation reaction as well as the o-(m)ethyl transfer reactions producing the tensyuic acids. Alternative approaches to increase their bulk in media extracts may aid in pathway validation. Going forward, a project to validate our proposed pathway will be required. This would include a systematic disruption of all clustered and non-clustered genes putatively involved in biosynthesis. The construction of strains overexpressing other genes in the alkyl citrate cluster could also be useful for NMR structural elucidation of compounds at low concentrations. Genes putatively involved in ω-oxidation and o-(m)ethylation would be prime candidates to overproduce carboxylated hexylitaconic acid and the tensyuic acids.

Also planned are strain constructions containing multiple gene deletions to generate citric acid through the disruption of the clustered *cha5* citrate synthase gene as well as the non-clustered citrate lyase subunit. Strains overproducing oxaloacetic acid and acetyl-CoA could also be developed by targeted deletion of the clustered FAS and *cha5* genes. Further afield, the introduction of enzymes from other pathways or non-native enzymes is planned to generate metabolites not known to be produced in *A. niger* or perhaps any other fungal species. The potential to produce itaconic acid, an industrial monomer used for producing plastics [160], may also be possible. Currently this monomer is produced industrially in *A. terreus* at concentrations of 80 g/L. Given that some acid producing strains of *A. niger* are capable of producing 200 g/L of citrate, an itaconate precursor, the potential exists to produce a strain of *A. niger* with theoretical yields of itaconate reaching 135 g/L. A key reason why this could be viable is due to the increased availability of citric acid in the cytosol for conversion to itaconate. This is one of the obstacles that need to be overcome in order to produce high titres of itaconate in *A. niger*.

Itaconate could then potentially be produced through the introduction of a citrate dehydratase and cis aconitate decarboxylase from *A. terreus* or by upregulation of what appears to be an *A. niger* native itaconate pathway (NRRL3_02448 – NRRL3_02450) to carry out the citrate to itaconate reactions [161].

References

1. Vining, L.C., *Functions of Secondary Metabolites.* Annual Review of Microbiology, 1990. **44**(1): p. 395-427.
2. Fox, E.M. and B.J. Howlett, *Secondary metabolism: regulation and role in fungal biology.* Current Opinion in Microbiology, 2008. **11**(6): p. 481-487.
3. Mao, S., et al., *A new polyacetylenic fatty acid and other secondary metabolites from the Chinese green alga Caulerpa racemosa (Caulerpaceae) and their chemotaxonomic significance.* Biochemical Systematics and Ecology, 2011. **39**(4-6): p. 253-257.
4. Rodrigues, A.P.D., et al., *A Novel Function for Kojic Acid, a Secondary Metabolite from Aspergillus Fungi, as Antileishmanial Agent.* PLoS ONE, 2014. **9**(3): p. e91259.
5. Vining, L.C., *Secondary metabolism, inventive evolution and biochemical diversity — a review.* Gene, 1992. **115**(1-2): p. 135-140.
6. O'connor, S.E., *Engineering of secondary metabolism.* Annu Rev Genet, 2015. **49**: p. 71-94.
7. Ligozzi, M., G. Lo Cascio, and R. Fontana, *vanA gene cluster in a vancomycin-resistant clinical isolate of Bacillus circulans.* Antimicrob Agents Chemother, 1998. **42**(8): p. 2055-2059.
8. Brakhage, A.A., et al., *Regulation of Penicillin Biosynthesis in Filamentous Fungi*, in *Molecular Biotechnolgy of Fungal beta-Lactam Antibiotics and Related Peptide Synthetases*. 2004, Springer Nature. p. 45-90.
9. Rathinasamy, K., et al., *Griseofulvin stabilizes microtubule dynamics, activates p53 and inhibits the proliferation of MCF-7 cells synergistically with vinblastine.* BMC Cancer, 2010. **10**(1).
10. Tobert, J.A., *Case history: Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors.* Nature Reviews Drug Discovery, 2003. **2**(7): p. 517-526.
11. Ogita, S., et al., *Metabolic engineering of caffeine production.* Plant Biotechnology, 2005. **22**(5): p. 461-468.
12. Vandenberghe, L.P.S., et al., *Microbial production of citric acid.* Braz. arch. biol. Technol., 1999. **42**(3): p. 22-29.
13. Shindia, A.A., et al., *Production of Gluconic Acid by Some Local Fungi.* Mycobiology, 2006. **34**(1): p. 22.
14. Stockmann-Juvala, H. and K. Savolainen, *A review of the toxic effects and mechanisms of action of fumonisin B1.* Human & Experimental Toxicology, 2008. **27**(11): p. 799-809.
15. Wang, J., et al., *Study of Malformin C, a Fungal Source Cyclic Pentapeptide, as an Anti-Cancer Drug.* PLOS ONE, 2015. **10**(11): p. e0140069.
16. Tsai, S.C., *A Fine Balancing Act of Type III Polyketide Synthase.* Chemistry & Biology, 2004. **11**(9): p. 1177-1178.
17. Korman, T.P., et al., *Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis.* Proceedings of the National Academy of Sciences, 2010. **107**(14): p. 6246-6251.
18. Felnagle, E.A., et al., *Nonribosomal Peptide Synthetases Involved in the Production of Medically Relevant Natural Products.* Molecular Pharmaceutics, 2008. **5**(2): p. 191-211.
19. Cseri, J., et al., *Analysis of the sensitizing effect of veratrum alkaloids to potassium on frog muscle.* Acta Physiol Acad Sci Hung, 1980. **56**(3): p. 289-301.
20. Marmulla, R. and J. Harder, *Microbial monoterpene transformationsâ€"a review.* Frontiers in Microbiology, 2014. **5**.
21. Schulze, B., P. Dąbrowska, and W. Boland, *Rapid Enzymatic Isomerization of 12-Oxophytodienoic Acid in the Gut of Lepidopteran Larvae.* ChemBioChem, 2007. **8**(2): p. 208-216.

22. Arnstein, H.R.V. and R. Bentley, *The biosynthesis of kojic acid. 1. Production from [1- 14 C] and [3:4- 14 C 2]glucose and [2- 14 C]-1:3-dihydroxyacetone.* Biochemical Journal, 1953. **54**(3): p. 493-508.
23. Ruiz, B., et al., *Production of microbial secondary metabolites: Regulation by the carbon source.* Critical Reviews in Microbiology, 2010. **36**(2): p. 146-167.
24. Osbourn, A., *Gene Clusters for Secondary Metabolic Pathways: An Emerging Theme in Plant Biology.* PLANT PHYSIOLOGY, 2010. **154**(2): p. 531-535.
25. Ansari, M., et al., *In silico analysis of methyltransferase domains involved in biosynthesis of secondary metabolites.* BMC Bioinformatics, 2008. **9**(1): p. 454.
26. Ichikawa, N., et al., *DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters.* Nucleic Acids Research, 2012. **41**(D1): p. D408-D414.
27. Inglis, D.O., et al., *Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of Aspergillus nidulans, A. fumigatus, A. niger and A. oryzae.* BMC Microbiology, 2013. **13**(1): p. 91.
28. Brown, D.W., T.H. Adams, and N.P. Keller, *Aspergillus has distinct fatty acid synthases for primary and secondary metabolism.* Proceedings of the National Academy of Sciences, 1996. **93**(25): p. 14873-14877.
29. Keller, N.P., G. Turner, and J.W. Bennett, *Fungal secondary metabolism — from biochemistry to genomics.* Nature Reviews Microbiology, 2005. **3**(12): p. 937-947.
30. Nierman, W.C., et al., *Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus.* Nature, 2005. **438**(7071): p. 1151-1156.
31. Lozano, M.J.F., *Characterization of Two Polyketide Methyltransferases Involved in the Biosynthesis of the Antitumor Drug Mithramycin by Streptomyces argillaceus.* Journal of Biological Chemistry, 2000. **275**(5): p. 3065-3074.
32. Gay, D.C., P.J. Spear, and A.T. Keatinge-Clay, *A Double-Hotdog with a New Trick: Structure and Mechanism of thetrans-Acyltransferase Polyketide Synthase Enoyl-isomerase.* ACS Chemical Biology, 2014. **9**(10): p. 2374-2381.
33. Pi, B., et al., *A Genomics Based Discovery of Secondary Metabolite Biosynthetic Gene Clusters in Aspergillus ustus.* PLOS ONE, 2015. **10**(2): p. e0116089.
34. Guo, C.J., et al., *Biosynthetic pathway for the epipolythiodioxopiperazine acetylaranotin in Aspergillus terreus revealed by genome-based deletion analysis.* J Am Chem Soc, 2013. **135**(19): p. 7205-7213.
35. Yu, J., et al., *Clustered Pathway Genes in Aflatoxin Biosynthesis.* Appl Environ Microbiol., 2004. **70**(3): p. 1253-1262.
36. Jin, J., et al., *Functional characterization and manipulation of the apicidin biosynthetic pathway in Fusarium semitectum.* Mol Microbiol., 2010. **76**(2): p. 456-466.
37. Bergmann, S., et al., *Genomics-driven discovery of PKS-NRPS hybrid metabolites from Aspergillus nidulans.* Nat Chem Biol, 2007. **3**(4): p. 213-217.
38. Abe, Y., et al., *Molecular cloning and characterization of an ML-236B (compactin) biosynthetic gene cluster in Penicillium citrinum.* Mol Genet Genomics., 2002. **257**(5): p. 636-646.
39. Chen, H., et al., *Molecular analysis of the cercosporin biosynthetic gene cluster in Cercospora nicotianae.* Mol Microbiol., 2007. **64**(3): p. 755-770.
40. Kato, N., et al., *Genetic safeguard against mycotoxin cyclopiazonic acid production in Aspergillus oryzae.* ChemBioChem, 2011. **12**(9): p. 1376-1382.
41. qiao, K., Y.H. Chooi, and Y. Tang, *Identification and engineering of the cytochalasin gene cluster from Aspergillus clavatus NRRL 1.* Metab Eng., 2011. **13**(6): p. 723-732.
42. Sims, J.W., et al., *Equisetin biosynthesis in Fusarium heterosporum.* Chemical Communications, 2005(2): p. 186.

43. Cacho, R.A., et al., *Identification and characterization of the echinocandin B biosynthetic gene cluster from Emericella rugulosa NRRL 11440.* J Am Chem Soc, 2012. **134**(40): p. 16781-16790.
44. Yin, W.B., et al., *A nonribosomal peptide synthetase-derived iron(III) complex from the pathogenic fungus Aspergillus fumigatus.* J Am Chem Soc., 2013. **135**(6): p. 2064-2067.
45. Chen, L., et al., *Genomics-driven discovery of the pneumocandin biosynthetic gene cluster in the fungus Glarea lozoyensis.* BMC Genomics, 2013. **14**(1): p. 339.
46. Maiya, S., et al., *Identification of a Hybrid PKS/NRPS Required for Pseurotin A Biosynthesis in the Human PathogenAspergillus fumigatus.* ChemBioChem, 2007. **8**(14): p. 1736-1743.
47. Awakawa, T., et al., *Pyranonigrin E: A PKS-NRPS Hybrid Metabolite from Aspergillus niger Identified by Genome Mining.* ChemBioChem, 2013. **14**(16): p. 2095-2099.
48. Ma, S.M. and Y. Tang, *Biochemical characterization of the minimal polyketide synthase domains in the lovastatin nonaketide synthase LovB.* FEBS Journal, 2007. **274**(11): p. 2854-2864.
49. Amnuaykanjanasin, A., et al., *Diversity of type I polyketide synthase genes in the wood-decay fungus Xylaria sp. BCC 1067.* FEMS Microbiology Letters, 2005. **251**(1): p. 125-136.
50. Gallo, A., M. Ferrara, and G. Perrone, *Phylogenetic Study of Polyketide Synthases and Nonribosomal Peptide Synthetases Involved in the Biosynthesis of Mycotoxins.* Toxins, 2013. **5**(4): p. 717-742.
51. Kehr, J., D. Gatte Picchi, and E. Dittmann, *Natural product biosyntheses in cyanobacteria: A treasure trove of unique enzymes.* Beilstein Journal of Organic Chemistry, 2011. **7**: p. 1622-1635.
52. Austin, M.B. and J.P. Noel, *The chalcone synthase superfamily of type III polyketide synthases.* Natural Product Reports, 2002. **20**(1): p. 79-110.
53. Yu, D., et al., *Type III polyketide synthases in natural product biosynthesis.* IUBMB Life, 2012. **64**(4): p. 285-295.
54. Kroken, S., et al., *Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes.* Proceedings of the National Academy of Sciences, 2003. **100**(26): p. 15670-15675.
55. Liu, L., et al., *Bioinformatical Analysis of the Sequences, Structures and Functions of Fungal Polyketide Synthase Product Template Domains.* Sci Rep, 2015. **5**(10463).
56. Beck, J., et al., *The multifunctional 6-methylsalicylic acid synthase gene of Penicillium patulum. Its gene structure relative to that of other polyketide synthases.* European Journal of Biochemistry, 1990. **192**(2): p. 487-498.
57. Schmitt, L., S. Kautz, and H.T. Lumbsch, *6-MSAS-like polyketide synthase genes occur in lichenized ascomycetes.* Mycol Res, 2008. **112**(2): p. 289-296.
58. Zabala, A.O., et al., *Fungal Polyketide Synthase Product Chain-Length Control by Partnering Thiohydrolase.* ACS Chem Biol, 2014. **9**(7): p. 1576-1586.
59. Morrow, G.W., *Bioorganic synthesis an introduction*. 2016, Oxford University Press: New York, NY. p. 235-236.
60. Hertweck, C., et al., *Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork* Nat. Prod. Rep., 2007. **24**: p. 162-190.
61. Lou, L., et al., *Biosynthesis of HSAF, a Tetramic Acid-Containing Macrolactam from Lysobacter enzymogenes.* Journal of the American Chemical Society, 2011. **133**(4): p. 643-645.
62. Khosla, C., et al., *Tolerance and specificity of polyketide synthases.* Annu Rev Biochem, 1999. **68**: p. 219-253.
63. Chiang, Y., et al., *Unraveling polyketide synthesis in members of the genus Aspergillus.* Applied Microbiology and Biotechnology, 2010. **86**(6): p. 1719-1736.
64. Tang, Y., S.C. Tsai, and C. Khosla, *Polyketide Chain Length Control by Chain Length Factor.* Journal of the American Chemical Society, 2003. **125**(42): p. 12708-12709.

65. Rasmus, J.N.F. *Animation of a minimal PKS*. [cited 2013 November 11]; Available from: www.rasmusfrandsen.dk/Animation%20of%20minimal%20PKS.ppt.
66. Ding, W., et al., *Biosynthetic investigation of phomopsins reveals a widespread pathway for ribosomal natural products in Ascomycetes.* Proc Natl Acad Sci U.S.A., 2016. **113**(13): p. 3521-3526.
67. Röttig, M., et al., *NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity.* Nucleic Acids Research, 2011. **39**(suppl): p. W362-W367.
68. Anke, T., D. Weber, and K. Esser, *The Mycota: A comprehensive treatise on fungi as experimental systems for basic and applied research*. XV ed. 2009, New York: Springer.
69. Evans, B.S., S.J. Robinson, and N.L. Kelleher, *Surveys of non-ribosomal peptide and polyketide assembly lines in fungi and prospects for their analysis in vitro and in vivo.* Fungal Genetics and Biology, 2011. **48**(1): p. 49-61.
70. Lombó, F., et al., *Deciphering the Biosynthesis Pathway of the Antitumor Thiocoraline from a Marine Actinomycete and Its Expression in Two Streptomyces Species.* ChemBioChem, 2006. **7**(2): p. 366-376.
71. Gross, H., et al., *The Genomisotopic Approach: A Systematic Method to Isolate Products of Orphan Biosynthetic Gene Clusters.* Chemistry & Biology, 2007. **14**(1): p. 53-63.
72. Marahiel, M.A., *Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis.* Journal of Peptide Science, 2009. **15**(12): p. 799-807.
73. *Non-Ribosomal Peptide Synthesis*. 2013 [cited 2016 August 22]; iGem competition website for 2013]. Available from: http://2013.igem.org/wiki/index.php?title=Team:Heidelberg/NRPS&oldid=353123.
74. Boettger, D. and C. Hertweck, *Molecular Diversity Sculpted by Fungal PKS-NRPS Hybrids.* ChemBioChem, 2012. **14**(1): p. 28-42.
75. Andersen, M.R., et al., *Comparative genomics of citric-acid-producing Aspergillus niger ATCC 1015 versus enzyme-producing CBS 513.88.* Genome Research, 2011. **21**(6): p. 885-897.
76. Wackler, B., et al., *Characterization of the Suillus grevillei Quinone Synthetase GreA Supports a Nonribosomal Code for Aromatic α-Keto Acids.* ChemBioChem, 2012. **13**(12): p. 1798-1804.
77. Schneider, P., S. Bouhired, and D. Hoffmeister, *Characterization of the atromentin biosynthesis genes and enzymes in the homobasidiomycete Tapinella panuoides.* Fungal Genetics and Biology, 2008. **45**(11): p. 1487-1496.
78. Helfrich, E.J.N. and J. Piel, *Biosynthesis of polyketides by trans-AT polyketide synthases.* Nat Prod Rep, 2016. **33**(231).
79. Chen, X., et al., *Iterative type I polyketide synthases involved in enediyne natural product biosynthesis.* IUBMB Life, 2014. **66**(9): p. 587-595.
80. Jenni, S., et al., *Structure of Fungal Fatty Acid Synthase and Implications for Iterative Substrate Shuttling.* Science, 2007. **316**(5822): p. 254-261.
81. Lynen, F., *On the structure of fatty acid synthase.* Eur J Biochem, 1980. **112**(3): p. 431-442.
82. Hopwood, D.A. and D.H. Sherman, *Molecular genetics of polyketides and its comparison to fatty acid biosynthesis.* Annu Rev Genet., 1990. **24**: p. 37-66.
83. Wallace, K.K., et al., *In vivo analysis of straight-chain and branched-chain fatty acid biosynthesis in three actinomycetes.* FEMS Microbiology Letters, 1995. **131**(2): p. 227-234.
84. Fujii, I., et al., *Identification of Claisen cyclase domain in fungal polyketide synthase WA, a naphthopyrone synthase of Aspergillus nidulans.* Chemistry & Biology, 2001. **8**(2): p. 189-197.
85. Metzger, U., et al., *The structure of dimethylallyl tryptophan synthase reveals a common architecture of aromatic prenyltransferases in fungi and bacteria.* Proceedings of the National Academy of Sciences, 2009. **106**(34): p. 14309-14314.

86. Cardoza, R.E., et al., *Identification of Loci and Functional Characterization of Trichothecene Biosynthesis Genes in Filamentous Fungi of the Genus Trichoderma.* Applied and Environmental Microbiology, 2011. **77**(14): p. 4867-4877.
87. Medema, M.H., et al., *antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences.* Nucleic Acids Research, 2011. **39**(suppl): p. W339-W346.
88. Khaldi, N., et al., *SMURF: Genomic mapping of fungal secondary metabolite clusters.* Fungal Genetics and Biology, 2010. **47**(9): p. 736-741.
89. Riley, D.R., et al., *Using Sybil for interactive comparative genomics of microbes on the web.* Bioinformatics, 2011. **28**(2): p. 160-166.
90. Brakhage, A.A., *Regulation of fungal secondary metabolism.* Nature Reviews Microbiology, 2013. **11**(1): p. 21-32.
91. Kotera, M., et al., *MUCHA: multiple chemical alignment algorithm to identify building block substructures of orphan secondary metabolites.* BMC Bioinformatics, 2011. **12**(Suppl 14): p. S1.
92. Isogai, A., et al., *Isolation and identification of (+)-hexylitaconic acid as a plant growth regulator.* Agricultural and Biological Chemistry, 1984. **48**(10): p. 2607-2609.
93. Hasegawa, Y., et al., *Tensyuic Acids, New Antibiotics Produced by Aspergillus niger FKI-2342.* CHEMICAL & PHARMACEUTICAL BULLETIN, 2007. **55**(9): p. 1338-1341.
94. Matsumaru, T., et al., *Synthesis and biological properties of tensyuic acids B, C, and E, and investigation of the optical purity of natural tensyuic acid B.* Tetrahedron, 2008. **64**(30-31): p. 7369-7377.
95. Newman, A.G. and C.A. Townsend, *Molecular Characterization of the Cercosporin Biosynthetic Pathway in the Fungal Plant Pathogen Cercospora nicotianae.* J Am Chem Soc, 2016. **138**(12): p. 4219–4228.
96. Lim, F.Y., et al., *Toward Awakening Cryptic Secondary Metabolite Gene Clusters in Filamentous Fungi*, in *Methods in Enzymology*. 2012, Elsevier BV. p. 303-324.
97. Girol, C.G., et al., *Regio- and Stereoselective Oxidative Phenol Coupling in Aspergillus niger.* Angewandte Chemie, 2012. **124**(39): p. 9926-9929.
98. Nielsen, K.F., et al., *Review of secondary metabolites and mycotoxins from the Aspergillus niger group.* Analytical and Bioanalytical Chemistry, 2009. **395**(5): p. 1225-1242.
99. Yang, X., et al., *Three Acyltetronic Acid Derivatives: Noncanonical Cryptic Polyketides from Aspergillus niger Identified by Genome Mining.* ChemBioChem, 2014. **15**(11): p. 1578-1583.
100. Zabala, A.O., et al., *Characterization of a Silent Azaphilone Gene Cluster from Aspergillus niger ATCC 1015 Reveals a Hydroxylation-Mediated Pyran-Ring Formation.* Chemistry & Biology, 2012. **19**(8): p. 1049-1059.
101. Liu, D., et al., *Nigerapyrones A–H, α-Pyrone Derivatives from the Marine Mangrove-Derived Endophytic Fungus Aspergillus niger MA-132.* J. Nat. Prod.,, 2011. **74**(8): p. 1787-1791.
102. Li, Y., et al., *Comparative Characterization of Fungal Anthracenone and Naphthacenedione Biosynthetic Pathways Reveals an α-Hydroxylation-Dependent Claisen-like Cyclization Catalyzed by a Dimanganese Thioesterase.* Journal of the American Chemical Society, 2011. **133**(39): p. 15773-15785.
103. Chiang, Y., et al., *Characterization of a polyketide synthase in Aspergillus niger whose product is a precursor for both dihydroxynaphthalene (DHN) melanin and naphtho-γ-pyrone.* Fungal Genetics and Biology, 2011. **48**(4): p. 430-437.
104. Tsai, H.F., et al., *Pentaketide melanin biosynthesis in Aspergillus fumigatus requires chain-length shortening of a heptaketide precursor.* J. Biol. Chem, 2001. **276**: p. 29292–29298.
105. Jørgensen, T.R., et al., *The molecular and genetic basis of conidial pigmentation in Aspergillus niger.* Fungal Genet Biol., 2011. **48**(5): p. 544-553.

106. Ehrlich, K.C. and B. Mack, *Comparison of Expression of Secondary Metabolite Biosynthesis Cluster Genes in Aspergillus flavus, A. parasiticus, and A. oryzae.* Toxins, 2014. **6**(6): p. 1916-1928.
107. Calvo, A.M., et al., *Relationship between Secondary Metabolism and Fungal Development.* Microbiology and Molecular Biology Reviews, 2002. **66**(3): p. 447-459.
108. Bergmann, S., et al., *Activation of a Silent Fungal Polyketide Biosynthesis Pathway through Regulatory Cross Talk with a Cryptic Nonribosomal Peptide Synthetase Gene Cluster.* Applied and Environmental Microbiology, 2010. **76**(24): p. 8143-8149.
109. Perrin, R.M., et al., *Transcriptional Regulation of Chemical Diversity in Aspergillus fumigatus by LaeA.* PLoS Pathogens, 2007. **3**(4): p. e50.
110. Lee, I., et al., *HdaA, a class 2 histone deacetylase of Aspergillus fumigatus, affects germination and secondary metabolite production.* Fungal Genetics and Biology, 2009. **46**(10): p. 782-790.
111. Anyaogu, D.C. and U.H. Mortensen, *Heterologous production of fungal secondary metabolites in Aspergilli.* Front Microbiol., 2015. **6**(77).
112. Curtis, R.W., *Curvatures and Malformations in Bean Plants Caused by Culture Filtrate of Aspergillus niger.* PLANT PHYSIOLOGY, 1958. **33**(1): p. 17-22.
113. Frisvad, J.C., et al., *Fumonisin and Ochratoxin Production in Industrial Aspergillus niger Strains.* PLoS ONE, 2011. **6**(8): p. e23496.
114. Bojja, R.S., et al., *Determining the Biosynthetic Sequence in the Early Steps of the Fumonisin Pathway by Use of Three Gene-Disruption Mutants ofFusarium verticillioides.* Journal of Agricultural and Food Chemistry, 2004. **52**(10): p. 2855-2860.
115. Clarke, T., *Drug companies snub antibiotics as pipeline threatens to run dry.* Nature, 2003. **425**(6955): p. 225-225.
116. Sukkar, E., *Why are there so few antibiotics in the research and development pipeline?* The Pharmaceutical Journal, 2013.
117. Mendelson, M. and M.P. Matsoso, *The World Health Organization Global Action Plan for antimicrobial resistance.* South African Medical Journal, 2015. **105**(5): p. 325.
118. Hota, B., et al., *Predictors of Clinical Virulence in Community-Onset Methicillin-Resistant Staphylococcus aureus Infections: The Importance of USA300 and Pneumonia.* Clinical Infectious Diseases, 2011. **53**(8): p. 757-765.
119. Ganter, B., et al. *Complacency kills. Antibiotic resistance still on the rise in Europe*. 2011 [cited 2016 October 21]; April 7:[Available from: http://www.euro.who.int/en/media-centre/sections/press-releases/2011/04/complacency-kills.-antibiotic-resistance-still-on-the-rise-in-europe.
120. McLeod, S.M., T.J. Dougherty, and M.J. Pucci, *Novel Antibacterial Targets/Identification of New Targets by Comparative Genomics*, in *Antibiotic Discovery and Development*. 2011, Springer Nature. p. 881-900.
121. Reen, F., et al., *The Sound of Silence: Activating Silent Biosynthetic Gene Clusters in Marine Microorganisms.* Marine Drugs, 2015. **13**(8): p. 4754-4783.
122. Proctor, R.H., et al., *Co-expression of 15 contiguous genes delineates a fumonisin biosynthetic gene cluster in Gibberella moniliformis.* Fungal Genetics and Biology, 2003. **38**(2): p. 237-249.
123. Pel, H.J., et al., *Genome sequencing and analysis of the versatile cell factory Aspergillus niger CBS 513.88.* Nat Biotech, 2007. **25**(2): p. 221-231.
124. Holm, D.K., et al., *Molecular and Chemical Characterization of the Biosynthesis of the 6-MSA-Derived Meroterpenoid Yanuthone D in Aspergillus niger.* Chemistry & Biology, 2014. **21**(4): p. 519-529.
125. Erickson, B., J.E. Nelson, and P. Winters, *Perspective on opportunities in industrial biotechnology in renewable chemicals.* Biotechnology Journal, 2011. **7**(2): p. 176-185.

126. Lee, J.W., et al., *Systems metabolic engineering of microorganisms for natural and non-natural chemicals.* Nature Chemical Biology, 2012. **8**(6): p. 536-546.
127. Lim, J., et al., *Solution structures of the acyl carrier protein domain from the highly reducing type I iterative polyketide synthase CalE8.* PLOS ONE, 2011. **6**.
128. Crawford, J.M., et al., *Starter unit specificity directs genome mining of polyketide synthase pathways in fungi.* Bioorganic Chemistry, 2008. **36**(1): p. 16-22.
129. Li, Y., W. Xu, and Y. Tang, *Classification, Prediction, and Verification of the Regioselectivity of Fungal Polyketide Synthase Product Template Domains.* Journal of Biological Chemistry, 2010. **285**(30): p. 22764-22773.
130. Master, E.R., et al., *A xyloglucan-specific family 12 glycosyl hydrolase from Aspergillus niger : recombinant expression, purification and characterization.* Biochemical Journal, 2008. **411**(1): p. 161-170.
131. Käfer, E., *Meiotic and Mitotic Recombination in Aspergillus and Its Chromosomal Aberrations*, in *Advances in Genetics Volume 19*. 1977, Elsevier BV. p. 33-131.
132. Storms, R., et al., *Plasmid vectors for protein production, gene expression and molecular manipulations in Aspergillus niger.* Plasmid, 2005. **53**(3): p. 191-204.
133. Aslanidis, C. and P.J. de Jong, *Ligation-independent cloning of PCR products (LIC-PCR).* Nucleic Acids Research, 1990. **18**(20): p. 6069-6074.
134. Semova, N., et al., *Generation, annotation, and analysis of an extensive Aspergillus niger EST collection.* BMC Microbiol, 2006. **6**(7).
135. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biology, 2010. **11**(10): p. R106.
136. Melamud, E., L. Vastag, and J.D. Rabinowitz, *Metabolomic Analysis and Visualization Engine for LC−MS Data.* Analytical Chemistry, 2010. **82**(23): p. 9818-9826.
137. Tsai, H.F., et al., *A developmentally regulated gene cluster involved in conidial pigment biosynthesis in Aspergillus fumigatus.* J Bacteriol. , 1999. **181**(20): p. 6469-6477.
138. Schrettl, M., et al., *Distinct roles for intra- and extracellular siderophores during Aspergillus fumigatus infection.* PLoS Pathogens, 2007. **3**(9): p. 1195-1207.
139. Khaldi, N. and K.H. Wolfe, *Evolutionary Origins of the Fumonisin Secondary Metabolite Gene Cluster inFusarium verticillioidesandAspergillus niger.* International Journal of Evolutionary Biology, 2011. **2011**: p. 1-7.
140. Keller, B.O., et al., *Interferences and contaminants encountered in modern mass spectrometry.* Analytica Chimica Acta, 2008. **627**(1): p. 71-81.
141. Bonnarme, P., et al., *Itaconate biosynthesis in Aspergillus terreus.* Journal of Bacteriology, 1995. **177**(12): p. 3573-3578.
142. Klitgaard, A., et al., *Aggressive dereplication using UHPLC–DAD–QTOF: screening extracts for up to 3000 fungal secondary metabolites.* Analytical and Bioanalytical Chemistry, 2014. **406**(7): p. 1933-1943.
143. Berg, J.M., J.L. Tymoczko, and L. Stryer, *Biochemistry*. 5 ed. 2002, New York: W H Freeman.
144. Gerike, U., et al., *Citrate synthase and 2-methylcitrate synthase: structural, functional and evolutionary relationships.* Microbiology, 1998. **144**(4): p. 929-935.
145. Mahlen, A., *Properties of 2-Decylcitrate Synthase from Penicillium spiculisporum Lehman.* European Journal of Biochemistry, 1971. **22**(1): p. 104-114.
146. Chen, H., et al., *Physiological characterization of ATP-citrate lyase in Aspergillus niger.* Journal of Industrial Microbiology & Biotechnology, 2014. **41**(4): p. 721-731.
147. Fujii, R., et al., *Biosynthetic Study on Antihypercholesterolemic Agent Phomoidride: General Biogenesis of Fungal Dimeric Anhydrides.* Organic Letters, 2015. **17**(22): p. 5658-5661.

148. Williams, K., et al., *Heterologous Production of Fungal Maleidrides Reveals the Cryptic Cyclization Involved in their Biosynthesis.* Angewandte Chemie International Edition, 2016. **55**(23): p. 6784-6788.
149. Szwalbe, A.J., et al., *Novel nonadride, heptadride and maleic acid metabolites from the byssochlamic acid producer Byssochlamys fulva IMI 40021 – an insight into the biosynthesis of maleidrides.* Chem. Commun., 2015. **51**(96): p. 17088-17091.
150. Hildebrandt, H., U. Höfker, and G. Fels. *Carboxylic Acid Derivatives*. [cited 2017 Feb 1]; Available from: http://www.chemgapedia.de/vsengine/vlu/vsc/en/ch/12/oc/vlu_organik/c_acid/carbons_und_derivate.vlu.html.
151. Tuckwell, D., D.W. Denning, and P. Bowyer, *A public resource for metabolic pathway mapping of Aspergillus fumigatus Af293.* Medical Mycology, 2011. **49**(S1): p. S114-S119.
152. Gibbons, J.G. and A. Rokas, *The function and evolution of the Aspergillus genome.* Trends in Microbiology, 2013. **21**(1): p. 14-22.
153. Susca, A., et al., *Variation in the fumonisin biosynthetic gene cluster in fumonisin-producing and nonproducing black aspergilli.* Fungal Genetics and Biology, 2014. **73**: p. 39-52.
154. Kohli, G.S., et al., *Evolutionary distinctiveness of fatty acid and polyketide synthesis in eukaryotes.* The ISME Journal, 2016. **10**(8): p. 1877-1890.
155. Rausch, C., *Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs).* Nucleic Acids Research, 2005. **33**(18): p. 5799-5808.
156. Bodanszky, M. and G.L. Stahl, *The Structure and Synthesis of Malformin A.* Proceedings of the National Academy of Sciences, 1974. **71**(7): p. 2791-2794.
157. Castegna, A., et al., *Identification and Functional Characterization of a Novel Mitochondrial Carrier for Citrate and Oxoglutarate in Saccharomyces cerevisiae.* Journal of Biological Chemistry, 2010. **285**(23): p. 17359-17370.
158. Zhang, J., et al., *Determination of the cytosolic NADPH/NADP ratio in Saccharomyces cerevisiae using shikimate dehydrogenase as sensor reaction.* Sci Rep, 2015. **5**: p. 12846.
159. Tang, Y., S.C. Tsai, and C. Khosla, *Polyketide Chain Length Control by Chain Length Factor.* J Am Chem Soc, 2003. **125**(42): p. 12708-12709.
160. Huang, X., et al., *Improving itaconic acid production through genetic engineering of an industrial Aspergillus terreus strain.* Microbial Cell Factories, 2014. **13**(1).
161. van der Straat, L., et al., *Expression of the Aspergillus terreus itaconic acid biosynthesis cluster in Aspergillus niger.* Microbial Cell Factories, 2014. **13**(1): p. 11.