Characterization and regulation of biosynthetic gene clusters in *Aspergillus niger*

Gregory Evdokias

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

May 2020

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Gregory Evdokias

Entitled: Characterization and regulation of biosynthetic gene clusters in *Aspergillus niger*

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. David Kwan

_____ Examiner

Dr. Laurent Potvin-Trottier

_____ Examiner

Dr. Brandon Findlay

_____ Supervisor

Dr. Isabelle Benoit Gelber

Approved by _____

Graduate Program Director, Robert Weladji

_____ 2020 _____

Dean of Faculty, Pascale Sicotte

ABSTRACT

Characterization and regulation of biosynthetic gene clusters in *Aspergillus niger*

Gregory Evdokias

Fungi produce a broad range of secondary metabolites with various bioactivities that have allowed them to be used as antibiotics and pharmaceutical drugs. The genes encoding secondary metabolites are commonly organized contiguously into biosynthetic gene clusters. The close spacing of functionally related genes facilitates the identification of multiple parts of a secondary metabolite pathway, its regulation, and provides the possibility of discovering novel bioactive compounds.

However, (i) the function of each genes within biosynthetic gene cluster is often not well defined, (ii) the regulation of gene clusters is very complex and, so far, remains mostly uncharacterized.

Herein, three *Aspergillus niger* gene clusters were investigated. The transcriptional regulation of two clusters thought to be involved in malformins production and one unknown cluster not associated with any metabolites, was studied. The overexpression of transcription factors and knocked out of backbone enzymes were done by gene replacement. The production of malformins and others secondary metabolites was assessed by comparative metabolomics and transcriptomics using mass spectrometry and RNA-sequencing, respectively.

Acknowledgements
With deepest gratitude I would like to thank all the people who made this work possible.

To Dr. Isabelle Benoit Gelber, for whom I am infinitely fortunate to have had as my supervisor, I extend my thanks. Your patience and encouragement empowered me to accomplish more and made me the envy of graduate students.

To Professor Adrian Tsang, thank you for providing a space for my research and for challenging me.

To my committee members Dr. David Kwan and Dr. Lauren Potvin Trottier who attentively followed the progress of my research and provided constructive feedback.

To my personal mentor Sylvester Palys who sat with me through experiments and walked me through how the lab conducts research, and to Dr. Thi Thanh My Pham who took over many of his duties after Sylvester left.

To Dr. Marcos DiFalco for his technical expertise in mass spectrometry that fed me the data vital to my thesis.

To Montserrat Mora Ochomogo for being a hardworking and organized lab assistant, doing a lot of legwork for the NRRL3_00036 cluster.

To Marie-Claude Moisan for collecting RNA, helping with experiments, and driving me home.

To Dr. Cameron Semper who coordinated with me in the creation of the transcription factor over expressing strains.

Thanks to Sandrine Marqueteau, Catherine Bourassa, Susan Sillaots, Tricia John, Victoria Aguilar, Dr., Letian Song and Dr. Jean-Paul Ouedraogo for who were always available to offer advice in optimizing my experiments.

To Veronica Trozzo, who was always happy to point me the right way in the lab. To Chy Lam, Maria Isabel Casanova, Stephanie Gelinas, Amrita Roka Magar for their support and supplies.

Contribution of authors Contributor	Statement
Gregory Evdokias	Wrote the thesis. Designed and assembled vectors relating to gene knockouts. Conducted *A. niger* transformations. Performed analysis on the data. Transformed two strains as part of a larger project (*NRRL3_7888* and *NRRL3_08316*) and confirmed correct insertion of an additional six transcription factors in novel strains (*NRRL3_00988*, *NRRL3_01335*, *NRRL3_02628*, *NRRL_03978*, *NRRL3_05457*, and *NRRL3_08195*).
Sylvester Palys	Defining the secondary metabolite gene clusters. Initial work pointing to the potential malformins regulator and pathway.
Marcos DiFalco	Mass spectrometry instrument operation. Strong input in mass spectrometry data processing.
Montserrat Mora Ochomogo	Growth and mass spectrometry analysis of *NRRL3_00042* overexpression strain.
Marie-Claude Moisan	Grew and collected RNA for submission for RNA sequencing.
Cameron Semper	Providing advice for experiments. Ran RT-PCR for the NRRL3_00036 cluster strains.

Table of Contents

List of abbreviations

A domain: Adenylation domain
AMP: Adenosine monophosphate
ATP: Adenosine triphosphate
C domain: Condensation domain
CoA: Coenzyme A
GlaA: Glucoamylase
gRNA: guide RNA
McrA: methyl coenzyme-M reductase A
NRP: Non-ribosomal peptide
NRPS : Non-ribosomal peptide synthetase
PCR : Polymerase chain reaction
PPTase: Phosphopantetheinyltranferase
pyrG: CTP synthase
T or PCP domain: Thiolation domain
TE domain: Thioesterase domain
TIC: Total ion chromatogram
TPM: Transcripts per kilobase million

Table of figures

Table of tables

Introduction

Purpose

This study was conducted in order to associate the appearance of metabolites with the upregulation of transcription factors and the disappearance of said metabolites with the removal of biosynthetic enzyme genes. In this way the regulation of secondary metabolite gene clusters and the metabolic pathways of the metabolites they produce may be better understood. It was hypothesized that the transcription factors of the *NRRL3_08969* and *NRRL3_00036* clusters, *NRRL3_08965.1* and *NRRL3_00042*, regulate their clusters and that their upregulation would upregulate those clusters, resulting in secondary metabolite production.

Fungi hold untapped potential for the discovery of novel, clinically relevant, compounds in the form of secondary metabolites. While we already have the technology to analyze thousands of metabolites by mass spectrometry the data does not necessarily inform us on the relevance of these compounds. Further, metabolite pathways which remain silent under laboratory conditions do not produce metabolites and so do not show up in mass spectrometry. Activating the expression of secondary metabolite gene clusters can result in the expression of novel metabolites. These novel secondary metabolites are more likely to have bioactive effects, as a consequence of their role in nature, than the waste products of primary metabolism. Their discovery also allows for their association to the genes that produce them. This not only widens our databases of enzymes but also gives us targets to increase the expression of these metabolites, facilitating their production and study. Similarly, understanding how these pathways are regulated can also allow metabolite expression to be enhanced, with the added benefit of revealing the conditions under which these metabolites are expressed in nature, pointing to their evolutionary purpose.

These insights can allow the pursuit of long-term goals such as: the elimination of toxic secondary metabolite producing genes from industrial strains, the characterization of novel secondary metabolites, the increase in production of clinically relevant secondary metabolites, the engineering of secondary metabolite pathways and enzymes to produce new secondary metabolite derivatives, and the repurposing of secondary metabolite enzymes for the modification of new substrates.

1.1 Secondary metabolites: definition and origin

Organisms produce primary metabolites, which are necessary for the survival of the organism, and secondary metabolites, which are not required for the organism's survival. Instead, they provide a fitness advantage under certain conditions; namely the presence of predators and competitors (1). Common primary metabolites include macromolecules like DNA and proteins, which are necessary for an organism's growth and development. Secondary metabolites on the other hand exhibit a wide variety of bioactive effects. Bacteria, fungi, and plants use secondary metabolites to communicate and compete with competitors in their environment (2). Fungi, for instance, are known to produce antibiotics like penicillin (3). These antibiotics, in affecting the fitness of competitors, reduces competition for nutrients in the environment (3). Like primary metabolite production, secondary metabolite production requires the intake of nutrients. An organism must therefore balance the needs of growth and development with the needs of competing and communicating with other organisms in the environment. This results in a selective pressure to regulate secondary metabolism separately from primary metabolism; a regulation that may be altered under the different selective conditions of the laboratory (1). This pressure for

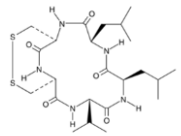
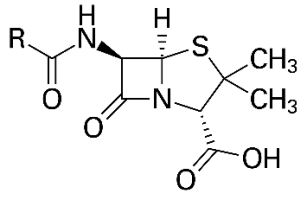
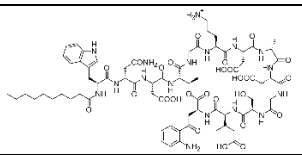
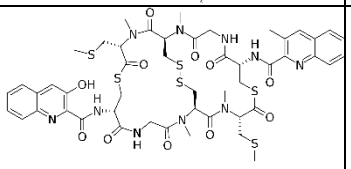
separate regulatory networks is one of the proposed reasons for co-localization of secondary metabolite genes into clusters (4).

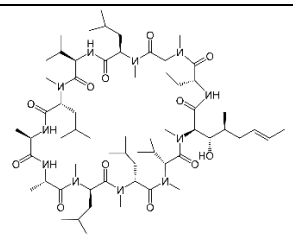
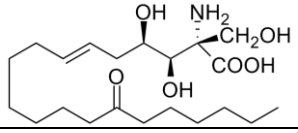
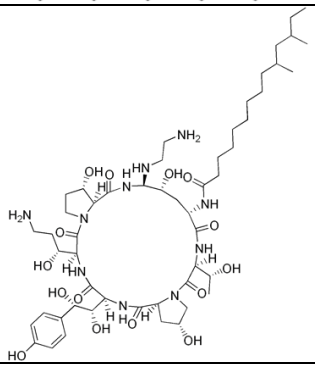
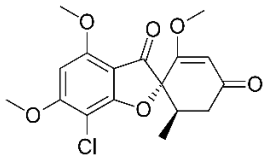
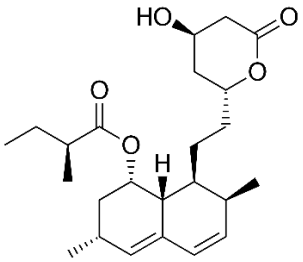
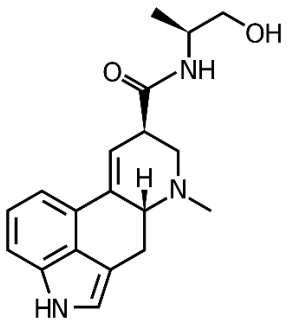
1.1.1 Fungal secondary metabolites structure and bioactive effects

Secondary metabolites are broadly categorized into different types according to their structure. These include, but are not limited to, non-ribosomal peptides, polyketides, fatty acids, alkaloids, terpenoids, and phenols. Secondary metabolites provide many avenues for research as well as many opportunities for the discovery of clinically relevant compounds.

This diverse array of secondary metabolites can produce a diverse array of bioactive effects. Some notable examples can be seen in table 1. Non-ribosomal peptides take up the bulk of this thesis and hold some unique characteristics which will be discussed. Polyketides are the most abundant secondary metabolite group, with structural diversity arising from principally from the building blocks acyl-CoA and malonyl-CoA (with subsequent modifications like methylation) (5). Alkaloids can hold a variety of interesting neurological effects (6), a famous example being caffeine which causes alertness. Humans have contended with alkaloids for our entire agricultural history as blighted crops caused mass poisonings with toxic alkaloids (7). Today, science has been able to isolate individual alkaloids and generate drugs, such as ergometrine, which had been for centuries used to help with childbirth, but has since, through isolation and proper dosage, has become safer to administer (8).

Table 1. Examples of clinically relevant secondary metabolites

Name	Structure	Class	Function	Species	References
Malformin C	 Malformin C	NRP	Plant-growth, anti-biotic, anti-viral, anti-cancer,	*Aspergillus niger, Aspergillus brasiliensis*	Anderegg *et al*. 1976(9), Theobald *et al*. 2018(10), Wang *et al*. 2015(11), Tan *et al*. 2015(12)
Penicillin		NRP	Antibiotic	*Penicillium species*	Künzler 2018(3)
Daptomycin		NRP	Antibiotic	*Streptomyces roseosporus*	Nguyen *et al*. 2006(13)
Thiocoraline		NRP	DNA binding: anti-cancer, anti-HIV, anti-bacterial	*Marine Micromonospora sp. L-13-ACM2-092*	Felnagle *et al*. 2008(14)

Cyclosporine		NRP	Immune suppression	*Tolypocladium inflatum*	Felnagle *et al*. 2008(14)
Myriocin		Fatty Acid	Immune suppression	*Mycelia sterilia*	Sasaki *et al*. 1994(15)
Caspofungin (an Echinocandin)		Lipopeptide	Antifungal	Derived from organisms *such as Aspergillus rugulovalvus*	Vardanyan *et al*. 2016(16)
Griseofulvin		Polyketide	Anti-fungal	*Penicillium griseofulvum*	Banani *et al*. 2016(17)
Lovastatin		Polyketide	Cholesterol reduction	*Aspergillus terreus*	Balraj *et al*. 2018(18)
Ergometrine		Alkaloid	Oxytocic	*Claviceps Purpurea*	De Groot *et al*. 1998(19)

1.1.2 Non-ribosomal peptides

Non-ribosomal peptides are a class of secondary metabolite. They are peptides that are not produced by the ribosome. As such, these peptides are not translated from mRNAs (20). Rather, they are assembled by enzymes from amino acids and are sometimes further modified by other

tailoring enzymes to meet the needs of the organism. The amino acids that are incorporated into these peptides can be of either D or L configuration. D and L amino acid configurations are enantiomers around the chiral carbon between the carboxylic acid group, the amino group, the variable group, and a hydrogen. L-amino acids are the amino acids typically incorporated into ribosomal peptides (21). These L-amino acids can have an R or S configuration, meaning a different handedness around the chiral center, depending on the amino acid. L-cysteine is (R)-cysteine while L-alanine is (S)-Alanine. This is why biochemistry uses the D and L convention. These peptides can be of varying lengths, including non-ribosomal peptides as large as 47 amino acids in the case of Polytheonamide B (22).

1.1.3 Defining the malformins non-ribosomal peptides

Malformins' effects were first observed in 1958 when a compound produced by *Aspergillus niger* caused malformations in the roots, stems and petioles of plants like the corn plant and bean plant (23). In the following years there were attempts to both characterize the compound's structure and bioactive effects. Multiple compounds sharing a similar structure were found in the *A. niger* derived extracts that caused the malformations. Named after the malformations they caused, these malformins were named alphanumerically as they were discovered. The discovery of malformin B, for instance, came after the discovery of malformin A. Both were isolated from different strains of *A. niger* (Malformin A from 56-39 and Malformin B from 56-30) (24). Since malformin B was shown to be two separate malformins by column chromatography, malformin B nomenclature was separated into malformin B1 and B2. Discovery of malformins proceeded in this way until the 11 named malformins listed in table 2 were discovered. Since these discoveries were made based off an incomplete understanding of malformins' molecular structures some redundancy was introduced in the naming convention. Molecules thought to be novel malformins were named before errors in the predicted structure were revealed. This resulted in malformin A3, B1b, and C having the same structure (25). The final count gives 9 malformin molecules produced by a single species with two different masses and very similar retention times.

1.1.4 Malformins' bioactive effects

Since its discovery, malformins have been included in drug screens and surveyed for a variety of bioactive effects. The first such study with malformin showed some level of inhibition in a few bacterial species such as *Bacillus subtilis* and *Escherichia coli*, with complete inhibition of bacterial growth at 2.5 and 5.0 mg/L malformin concentration respectively (26). In addition to plant malformations malformins have also been observed to stimulate plant growth additively with gibberellin in the stems of mung bean plants (27). Studies of malformins anti-cancer activities have also been conducted, with some success. While numbers, of course, vary between cell type and organism, one study reported Malformin C had an IC_{50} of 0.18 ± 0.023μM in HCT 116 cells (cancer cells), equating to a LC_{50} to IC_{50} of 4:1 (11). This narrow window could limit its clinical application, especially if the dose-response curve limits the margin of safety. It is possible that further alterations to malformins structure could reduce its toxicity or increase its effectiveness, widening its therapeutic window. Wang *et al*., for instance, plan on synthesizing new analogs of malformin C to test their effects (11). Even within known malformins molecules there is evidence of different molecules exhibiting different toxicities, with one study reporting that malformin C is more toxic in mammals than malformin A1 (LD_{50} of 0.9 mg/kg vs 3.1 mg/kg intraperitoneally, respectively) (28). Malformin A1 has shown anti-viral activity at an IC_{50} of 37.2 μM by lesion assay, while the antiviral positive control ningnanmycin gave an IC_{50} of 154.9 μM (12).

1.1.5 Malformins structures

Malformins are nine different non-ribosomal peptides composed of five amino acids each. The core structure is composed of two D-cysteine linked sequentially, followed by an L-Valine. The remaining amino acids are a D amino acid of either leucine, Isoleucine, or Valine, followed by an L amino acid of either leucine, isoleucine, valine, or alloisoleucine. The last amino acid is bonded back to the first cysteine by a peptide bond, forming a cyclic peptide. Both cysteines are also linked by a disulfide bond. These various malformins are listed in table 2 with their amino acid sequence.

Malformins has been observed in several *Aspergillus* species, though the exact malformins present have not always been identified. In addition to *A. niger* and *Aspergillus brasiliensis* malformins has been observed in *Aspergillus welwitschiae*, *Aspergillus awamori*, *Aspergillus ficuum*, *Aspergillus phoenicis*, and *Aspergillus tubingensis* (29–31). Several other species, such *Aspergillus luchensis*, also have homologous gene clusters that may produce malformins (10).

Table 2. Various forms of malformin and their amino acid sequence

Malformin	Amino Acid Sequence	Monoisotopic mass	Produced synthetically
Malformin A1	D-Cys-D-Cys-L-Val-D-Leu-L-Ile	529.239	Yes
Malformin A2	D-Cys-D-Cys-L-Val-D-Leu-L-Val	515.224	Yes
Malformin A3	D-Cys-D-Cys-L-Val-D-Leu-L-Leu	529.239	Yes
Malformin A4	D-Cys-D-Cys-L-Val-D-Ile-L-Val	515.224	No
Malformin B1a	D-Cys-D-Cys-L-Val-D-Leu-L-alloIle	529.239	No
Malformin B1b	D-Cys-D-Cys-L-Val-D-Leu-L-Leu	529.239	Yes
Malformin B2	D-Cys-D-Cys-L-Val-D-Val-L-Leu	515.224	Yes
Malformin B3	D-Cys-D-Cys-L-Val-D-Ile-L-Leu	529.239	No
Malformin B4	D-Cys-D-Cys-L-Val-D-Ile-L-Ile	529.239	No
Malformin B5	D-Cys-D-Cys-L-Val-D-Val-L-Ile	515.224	No
Malformin C	D-Cys-D-Cys-L-Val-D-Leu-L-Leu	529.239	Yes

1.2 Non-ribosomal peptide synthetases

Non-ribosomal peptides are produced by non-ribosomal peptide synthetases (NRPSs). NRPSs have three core domains that are required for the enzyme to make a peptide. The adenylation (A) domains binds a specific kind of amino acid and charges it with the addition of an adenosine mono phosphate (AMP) molecule to the carboxylic acid domain by adenosine triphosphate (ATP) hydrolysis. This use of ATP is why NRPSs are considered synthetases instead of synthases. The thiolation (T or PCP) domains hold the growing peptide chain by binding a sulfur atom to the location the AMP was previously. The condensation (C) domain binds the amino acid to the growing peptide (32). Proteins that only contain some subset of these three domains are often annotated as NRPS-like. The mechanism by which these domains assemble the non-ribosomal

peptide can be seen in figure 1. A fourth domain, the thioesterase (TE) domain, is sometimes responsible for release of the molecule, often through cyclization of the peptide (33). In addition to these there are optional domains which can be found within the NRPS enzyme. The optional domains include methyl transferases, reductases, epimerases, and the aforementioned thioesterase domains.

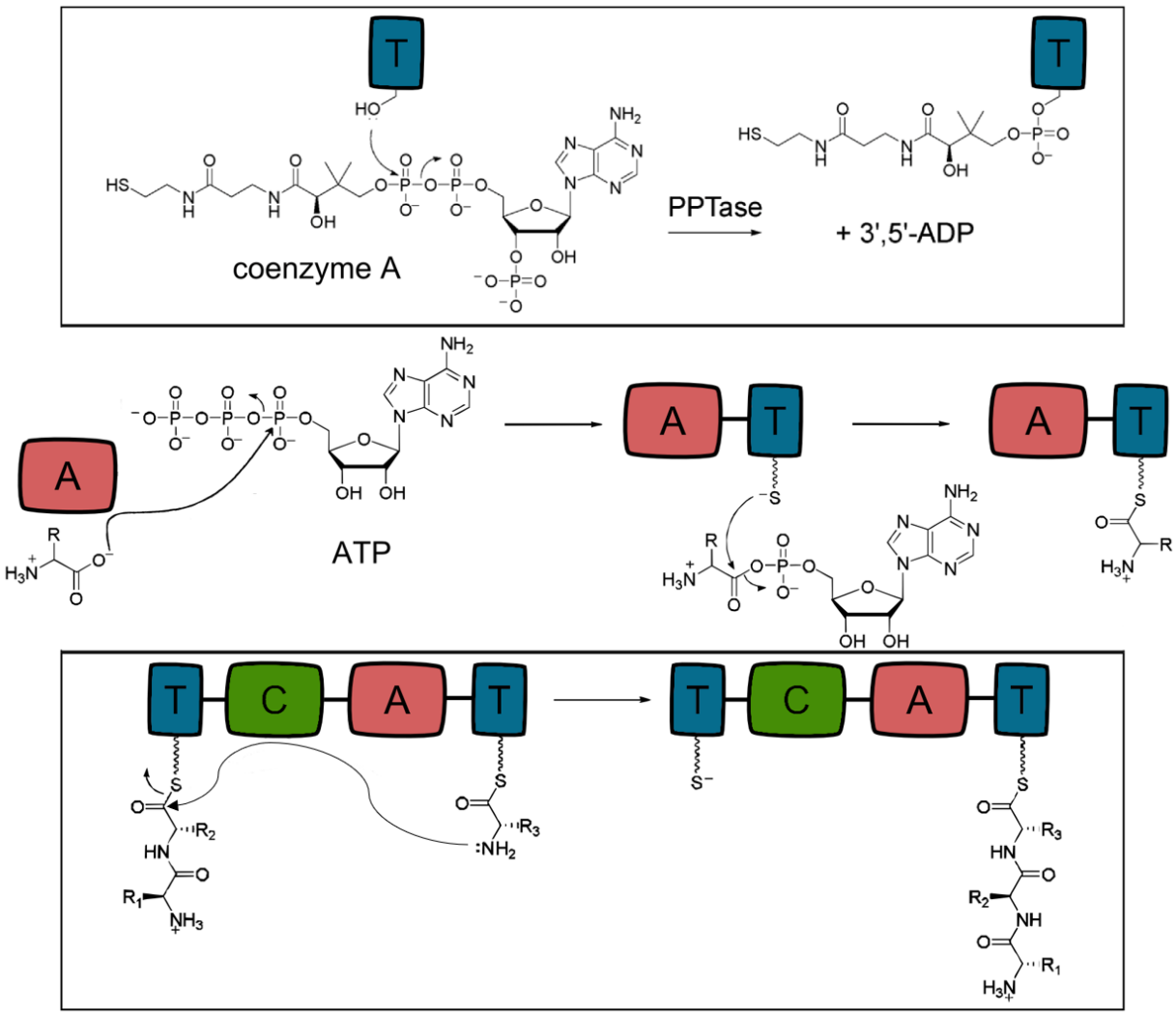


Figure 1. NRPS reaction mechanism, from activation of the T domain with coenzyme A to the addition of a second amino acid. The T domain must be activated by binding of phosphopantetheine from coenzyme A to a serine residue using a separate enzyme, phosphopantetheinyltranferase (PPTase). The A domain then charges its respective amino acid with an adenosine triphosphate, allowing it to bind to the thiol residue of phosphopantetheine. When two T domains are charged with an amino acid the C domain can catalyze the addition of an amino acid from one to the other. This figure was modified from 2013.igem.org/wiki/index.php?title=Team:Heidelberg/NRPS&oldid=353123 (34).

Together the core domains of the NRPS are called tridomains and are organized in units of A-C-T. Amino acids are added sequentially for each tridomain in sequence to build the peptide. Since

the peptide sequence lines up with the tridomain sequence we say they are co-linear. This cannon is often subverted, however, as in many NRPSs we see repeats and out of order domains. This leads to the classification of NRPS into three categories: Linear, iterative, and non-linear. Linear NRPSs are the simplest, working sequentially one tridomain after another. For this reason, they are the design currently used for synthetic NRPSs. Iterative NRPSs reuse domains and can feature multiple C or T domains for a given A domain. In non-linear NRPSs domain order is as of yet unpredictable. T or C domains can be lacking for a given A domain, or there could be multiple T or C domains for a given A domain (35).

Each A domain is said to have its own specificity conferring code. This means that any given A domain charges only a single type of amino acid. Prediction programs, however, are not much better than chance at predicting which amino acid an uncharacterized A domain is specific towards (36). Furthermore, these domains are not easily swappable. Doing so often does not produce a functional protein. Attempts at making synthetic NRPSs involve selecting tridomains from NRPSs that have the same specificity in the downstream tridomain in the native protein as the designed protein. This is because the links between domains have some degree of conservation, disrupting the folding if they are not next to the appropriate features (37). A more successful example of engineering these tri domains in the combinatorial biosynthesis of new compounds is with the Daptomycin NRPS, which resulted in the generation of novel antibiotics (13). This highlights the need to characterize more NRPS enzymes in order to better model NRP synthesis, or at least generate as many tridomain combinations as possible. NRPS enzymes are capable of producing multiple products, with recombined NRPSs losing some product diversity even when coded for the same amino acids, indicating more complexity in amino acid selection than just the A domains (37).

1.3 Biosynthetic gene clusters

Gene clusters are adjacent genes with some factor relating them to one another, though definitions of this relating factor are not always consistent. This factor could be common function, co-regulation of expression, or greater proximity of genes than expected by chance when viewed across organisms. Clustering of genes is present across the tree of life, though in higher forms of life with longer generation times, more complicated regulation, and less horizontal gene transfer they are less common. The clusters are commonly found in fungi and bacteria, with fungi typically carrying more biosynthetic gene clusters in their larger, more complicated genomes, than bacteria. Often these gene clusters code for various proteins facilitating secondary metabolite production and use (38). In these cases, clusters can be defined starting with the backbone enzyme, such as an NRPS, which produces the core of the molecule and working outwards toward relevant or relatively closely spaced genes. These other genes could be tailoring enzymes, such as reductases, which modify the metabolite structure, transporters, which export the metabolite extracellularly, and fungal specific transcription factors, evolved specifically in fungi to regulate fungal systems like gene clusters. The function of any given gene to secondary metabolite production is not always immediately apparent, and it is possible for an unrelated gene to find itself inserted into a gene cluster. It is also possible for a seemingly relevant gene to be included in a cluster's annotation without having a role in metabolite production or without being similarly regulated. Experiments need to be conducted to establish whether the role of the gene fits the cluster definition that was used to assign it to the cluster. In this way the study of gene clusters also gives us insights into their true boundaries. A rough diagram of the *NRRL3_08969* gene cluster can be seen in figure 2.

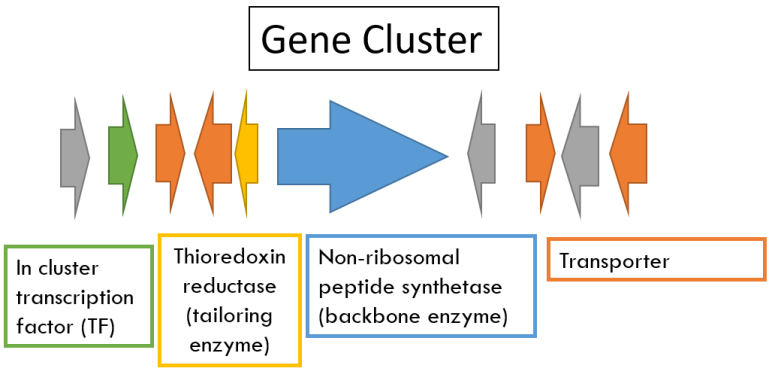


Figure 2. Diagram of the NRRL3_08969 gene cluster, which includes the basic features found in many gene clusters: adjacent genes, a backbone enzyme, a tailoring enzyme, transporters, a transcription factor, all with different gene orientations and lengths.

Having a cluster under the regulatory control of a regulatory system separate from housekeeping functions allows many of these clusters to be cryptic, or in other words transcriptionally silent (39). The metabolites these clusters may be capable of producing may not be seen under standard laboratory conditions. In these cases, the clusters must be induced or in some way upregulated to detect metabolites. In other cases, these clusters may produce known compounds where the link between the two has not been experimentally established.

1.4 Fungal secondary metabolite regulation

Genes are regulated by a complex interplay of molecules which facilitate and block access of transcriptional machinery. Proteins which bind specific DNA sequences in order to affect the level of transcription are called transcription factors (40). Fungi have evolved their own transcription factors called fungal-specific transcription factors, which belong to gene families not found in other kingdoms of life (40).

1.4.1 Master regulators

Master regulators are transcription factors that regulate an unusually large number of genes.

McrA (*NRRL3_03076*) is a transcription factor and negative regulator of at least 10 secondary metabolite clusters in *A. nidulans*, as well as the regulation of hundreds of individual genes (41). Deletion of the negative regulator has led to upregulation of several secondary metabolites, such as sterigmatocystin and nidulanin A, as well as the discovery of two new compounds. Overexpression of *McrA* showed reduced secondary metabolite expression. Expression analysis revealed expression of 112 otherwise silent genes, though levels remained low. McrA is a 391 amino acid protein containing a Zn(2)-C6 fungal-type DNA-binding domain at approximately the 138-204 amino acid position, according to InterPro scan.

LaeA is a master regulator of secondary metabolites in *Aspergilli*, influencing the expression of 9.8% of genes, 4.3% downregulated and 5.5% upregulated, in *Aspergillus fumigatus* (42). Of secondary metabolite genes, 97% showed downregulation in the *LaeA* knockout (42). LaeA is also connected to a light sensitive response via the velvet complex, which causes secondary metabolism inhibition in response to light (43). LaeA is 375 amino acids in length with an S-adenosyl-L-methionine-dependent methyltransferase domain running from amino acid 84-314, according to

InterPro scan. The exact mechanism of action and substrate of LaeA is not clear, besides its ability to self-methylate (44).

1.4.2 Epigenetic Regulation

Epigenetics are the study of gene expression due to factors outside of the genetic code itself. A large part of epigenetics is the control of expression through chromatin states. In cells, DNA is found wrapped around proteins called histones. The tightness of this association affects how accessible DNA is to transcription machinery such as transcription factors and RNA polymerase. When DNA is tightly wrapped around histones the DNA protein complex is called heterochromatin and transcription is reduced. Conversely, when the association is loose, the complex is called euchromatin and genes are accessible.

The state of chromatin is dependent on a number of different factors, including cell fate, cell cycle position, and the nature of nearby promoters. Given the right context, factors are brought in to modify histones and open up the chromatin. Different transcription factors, for instance, have varying capacities to recruit remodeling complexes. How far along the DNA the chromatin remodeling extends depends on the distribution of chromatin remodeling nucleation sites and the presence of certain barriers, called "barrier insulators" or "boundary elements" (45). These elements create a balance between maintenance of euchromatin and heterochromatin spreading.

The remodeling of chromatin states is accomplished by modifying histones with acetylation, methylation, sumoylation, ADP ribosylation, phosphorylation, deamination, and ubiquitylation (46). These modifications can later the strength of the binding between DNA and histones. DNA methylation and demethylation biproducts can also alter the strength of this interaction (47).

Many biosynthetic gene clusters are found in transcriptionally silent sub-telomeric regions of fungal genomes. In *Aspergillus nidulans*, for instance, deletion of an epigenetic regulator responsible for 3K4 methyltransferase activity, *cclA*, resulted in production of several secondary metabolites not before seen in the species (48).

The presence of a gene cluster transcription factor may not cause an optimal level of cluster expression if the cluster remains in a silent region of the genome. Introducing a strong promoter to the cluster may recruit proteins like the HAP complex which could create a euchromatin island extending to neighboring genes.

Given that it is possible for chromatin opening to propagate itself, and that clusters are thought to be regulated as a whole, it is possible that introducing sequences that assist chromatin opening to a cluster could help not only increase the expression of target genes, but nearby genes as well. There is evidence that epigenetic regulation can regulate metabolite clusters as a whole without affecting nearby genes out of cluster.

1.4.3 Secondary metabolite production induction

Fungi have evolved to control secondary metabolite expression to limit their production when the metabolites are not useful. In this way they can preserve resources for growth and reproduction. Many different strategies have evolved by which fungi detect the correct conditions for production of any given metabolite. Some respond to global regulators sensitive to iron (CCAAT binding complex) or nitrogen depletion (AreA), light (velvet complex), and pH (PacC) (49). Some metabolites have even been observed in response to radiation and zero-gravity, though experiments involving these factors would require more specialized equipment (50). Secondary metabolite gene cluster specific regulators can respond to more specific stimuli, such as molecules produced from other organisms. A more complex example of secondary metabolite production involving

interaction between different species are loline alkaloids, which requires enzymes from both *Epichloë siegelii* and some of its plant hosts (51). While some strategies are more complicated than others induction of secondary metabolite production by external stimuli can reveal novel metabolites and their mechanisms or regulation.

1.5 *Aspergillus niger*

Due to the ability of *Aspergillus niger* to withstand a wide range of temperature and humidity, and its ability to grow on many substrates, it can be found throughout the world in many climates (52). *A. niger* has been found in polar, desert, water-related, saltern, agricultural, and endozoic environments (52). *Aspergillus niger* is a mesophilic and saprophytic (52). It is a common contaminant in human food, appearing even in cereals due to its tolerance to low water activity (52).

A. niger is used in a variety of applications, with different strains being adopted towards different ends. Industrially, ATCC 1015 is used for citric acid production and CBS513.88 is used for enzyme production (53). In research laboratories, strains like NRRL 3 have been used for research (53). These strains have been used to create transformational tools to optimize production, as well as study genomics, proteomics, and metabolomics (53). Towards this end expression vectors with selectable markers have been developed. Further, a manually curated NRRL 3 genome has been made available (54). In this study the strain NRRL 2270 was used as a parental strain for genetic manipulation. This allows the NRRL 3 genome to be used since there are only 34 single nucleotide polymorphisms between NRRL 3 and ATCC 1015. NRRL 2270 is a spontaneous mutant of the aforementioned citric acid producing ATCC 1015 strain (53).

One tool developed in *A. niger* is the use of the glucoamylase promoter to inducibly upregulate genes. The glucoamylase promoter is the promoter of the glucoamylase gene, which is an enzyme that digests starches into glucose subunits. It is commonly used in *A. niger* genetics to express recombinant genes due to the promoter's ability to drive high levels of expression in the presence of maltose. Deletion analysis has shown that only 214 bp of this promoter is required for transcription, but also that high levels of expression require two sequences at -489 to -414 bp and -390 to -345 bp upstream of the start codon (55, 56). In *A. niger* a two subunit protein named AngCP was observed to bind these two CCAAT regions (57). AngCP, also known as AngHAPC (*Aspergillus niger* HAPC), is the homolog of the yeast HAP3 protein and *Aspergillus nidulans* AnCF protein (56). These all act as components to their respective HAP complexes, which act as transcriptional enhancer and remodel chromatin structure (58).

A. niger is an industrial strain for citric acid synthesis and a model organism for genetic manipulation. This means there are robust tools available to investigate these genes, and potential downstream applications.

1.5.1 *Aspergillus niger* gene clusters

Secondary metabolite genes often belong to gene clusters. This can simplify the process of gene discovery as relevant genes can sometimes be found nearby and expressed under the same regulatory controls as one another.

In *A. niger* there are 84 bioinformatically predicted gene clusters (59). This is compared to 39 predicted in *A. fumigatus*, 71 for *Aspergillus nidulans*, and 75 for *Aspergillus oryzae,* though the same paper found 81 gene clusters in *A. niger* so there is still the possibility of gene cluster discovery in all these species (60). Of these 84 gene clusters 60 contain transcription factors, and of those 60 transcription factors 58 have been over-expressed (manuscript in progress).

NRPSs are one of the two most common classes of secondary metabolites in fungi, the other being polyketides (20). In *Aspergillus niger* (*A. niger*) alone there are 18 NRPS open reading frames, 21 putative NRPS-like open reading frames, and 9 putative hybrid NRPS/PKS open reading frames (59). These putative genes have the potential to produce therapeutically relevant metabolites, or toxic compounds.

Nine of the 18 NRPSs in *A. niger* are within clusters containing a nearby transcription factor. This is also the case with 13 of the 20 NRPS-like open reading frames and 4 of the 9 hybrid NRPS/PKS open reading frames. All these open reading frames belong to putative clusters according to in-lab definitions where we start from these backbone enzymes and work outwards looking for relevant genes. What this means is there is still more work to be done to determine how and if these clusters of genes cooperate. Some of these may not be true clusters in any functional sense.

The products of six *A. niger* gene clusters have been published. These, along with another unpublished cluster, can be seen listed in table 3.

Table 3. Characterized secondary metabolite gene cluster in *A. niger*

Cluster (Backbone gene)	Class of metabolite	Transcription factor	Product	References
NRRL3_00147	PK	NRRL3_00148	Azanigerone (D, E, B)	Zabala *et al*. 2012 (61)
NRRL3_02189	PK	NRRL3_02186	Fumonisin (B2, B1, B4)	Aerts *et al*. 2017 (62)
NRRL3_06291	PK	NRRL3_06287	Yanuthone D	Holm *et al*. 2014 (63)
NRRL3_09549	PK	NRRL3_09549	BMS (192548, 2)	Li *et al*. 2011 (64)
NRRL3_10128	NRP/PKS Hybrid	NRRL3_10124	Pyranonigrin E	Awakawa *et al*. 2013 (65)
NRRL3_11031	NRP/PKS Hybrid	NRRL3_11029a	Carlosic acid	Yang *et al*. 2014 (66)
NRRL3_11763/ NRRL3_11767	Fatty acid	NRRL3_11765	Alkylcitric acid	Palys *et al*. Unpublished (67)

Ongoing research is being conducted as part of a larger project where gene cluster transcription factors in the *A.niger* NRRL 2270 strain are upregulated. The goal is to determine which of these regulators upregulate their cluster and what kinds of compounds are produced from that upregulation. The results of this thesis involve three of these transcription factors.

Results

This project contains two parts: One where the cryptic gene cluster *NRRL3_00036* was investigated and another where the regulation of malformins production was investigated. As a general goal the role of gene cluster transcription factors in regulating their cluster was examined.

Overexpression strains were generated using two vectors. One, ANEp8, contained the Cas9 gene, a *pyrG* auxotrophic marker, and a guide RNA that targeted the glucoamylase locus. The other, pJETC, contained a template for homologous repair. Transformations were done into a Δ*KusA*Δ*PyrG* NRRL 2270 *A. niger* strain incapable of undergoing non-homologous end joining (68). This was done to better select for colonies where homologous repair occurred as no colonies will have undergone non-homologous end joining. The empty vector controls are colonies taken after transformation with an ANEp8 plasmid not containing a guide RNA and absent a homologous repair template. The results of three strains overexpressing gene cluster transcription factors *NRRL3_00042*, *NRRL3_08965.1*, and *NRRL3_07873* in the glucoamylase locus are presented in this thesis.

Prior to this project multiple random integration strains were generated for the *NRRL3_07873* transcription factor gene. This was done by introducing a plasmid (ANIp7) with a selective marker (pyrG), transcription factor gene with glucoamylase promoter, and lacking an origin of replication. In order to survive the strains were required to incorporate the plasmid into their genome. These strains were quality controlled for integration, but the number of integration events and location of integration was not determined. For this project, a new random integration strain was generated.

To further characterize the *NRRL3_00036* gene cluster the *NRRL3_00036* backbone enzyme gene was knocked out in the *NRRL3_00042* overexpression background. Three independently transformed colonies were taken from this transformation: *NRRL3_00036* knockout colony 3, *NRRL3_00036* knockout colony 5, and *NRRL3_00036* knockout colony 10.

An *NRRL3_03076* (*McrA*) depletion strain was generated in the Δ*KusA*Δ*PyrG* NRRL 2270 background to characterize its metabolome and look for involvement in malformins production.

2.1 *NRRL3_00036* Cluster

Separate from malformins production, the cryptic *NRRL3_00036* cluster was investigated with the purpose of discovering novel metabolites.

NRRL3_00036 is one of the 18 *A. niger* NRPSs. According to manual annotations (54), this cluster contains a single transcription factor, *NRRL_00042*. In addition to the backbone enzyme are several genes that could be involved in modifying metabolite backbones, such as cytochrome p450 and a short-chain dehydrogenase/reductase. Cluster genes can be seen in table 4.

Table 4. Secondary metabolite gene cluster in which the backbone enzyme NRRL3_00036 resides

Gene Number	Gene Annotation
NRRL3_00036	**non-ribosomal peptide synthetase**
NRRL3_00037	FAD-binding domain-containing protein
NRRL3_00038	hypothetical protein
NRRL3_00039	MFS-type transporter
NRRL3_00040	cytochrome P450
NRRL3_00041	NAD(P)-binding domain-containing protein
NRRL3_00042	**fungal-specific transcription factor**
NRRL3_00043	FAD-binding domain-containing protein
NRRL3_00044	NmrA-like family protein
NRRL3_00046	hypothetical protein
NRRL3_00048	short-chain dehydrogenase/reductase

Gene annotations have been manually curated (54). Currently, to our knowledge, there are no publications discussing the function of the *NRRL3_00036* gene in *A. niger* or orthologous genes in other *Aspergillus* species. It is annotated as an NRPS in UniProt, NCBI, and within the annotated Center for Structural and Functional Genomic database.

2.1.1 Unique compounds associated with the *NRRL3_00036* cluster

The *NRRL3_00042* overexpression strain, being already available for testing as part of the larger transcription factor project, was analyzed for production of novel metabolites. This would indicate regulatory control of these metabolites by the *NRRL3_00042* transcription factor. Given that NRRL3_00042 is expected to regulate the cluster in which it resides the clusters backbone enzyme, NRRL3_00036, is expected to be responsible for these novel metabolites. A knockout of the *NRRL3_00036* gene was created in order to establish this relation.

The knockout of *NRRL3_00036* was conducted by making a double strand cut after the 211th base pair in the coding region of the gene and repairing it with 60 base pair and 90 base pair homologous repair templates omitting the genomic sequence between 354 base pairs upstream of the gene and 317 base pairs downstream of the gene. Five peaks, large enough to see on the total ion chromatogram, were observed to disappear in the *NRRL3_00036* knockout. These retention times peaks can be seen in figure 3.

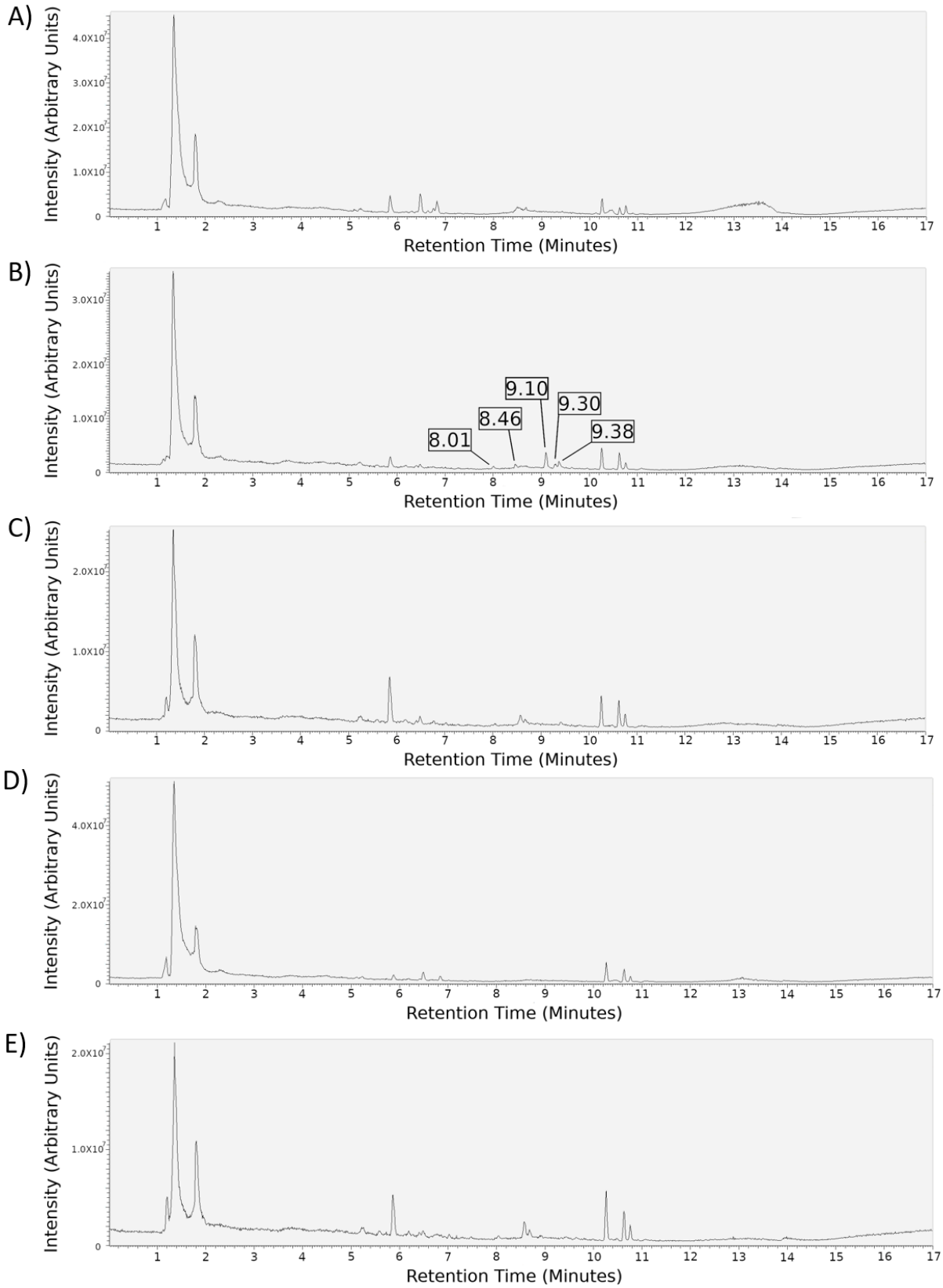


Figure 3. Total ion chromatogram of: A) empty vector control, B) *NRRL3_00042* overexpression strain, C) the third colony from a *NRRL3_00036* knockout transformation, D) the

fifth colony from a *NRRL3_00036* knockout transformation, E) the 10th colony from a *NRRL3_00036* knockout transformation.

The 9.10 retention time peak is the largest and contains a large mass peak at 425.1368 (Peak intensity across strains shown in figure 4). This peak shows two isotopic sub-peaks. There is a potential ammonium and sodium adduct peak at 442.1599 and 447.1151, though these potential adduct peaks have no isotopic sub-peaks and are relatively small at $1.6X10^4$. These mass peaks can be seen in figure 5.

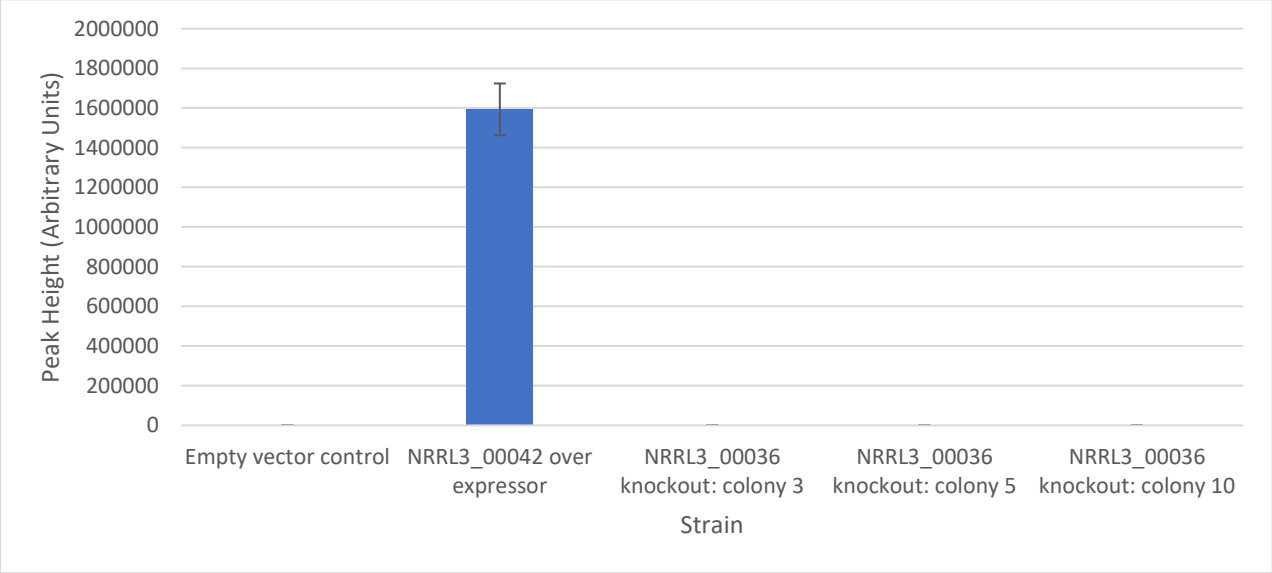


Figure 4. Peak intensity of mass 425.1368 at retention time 9.10.

Strains were grown in triplicate on 190µL MMJ for 6 days in 96 well plates at 30˚C. The 425.1368 mass peak at a retention time of 9.10 minutes is the highest unique peak seen only in the *NRRL3_00042* overexpression strain, and not in the empty vector control or the three independent *NRRL3_00036* deletion mutants. Values for mass spectrometry are reported as arbitrary units as mass spectrometry does not accurately represent amount of compound due to the variable sensitivity of mass spectrometry systems towards different compounds (69). The mass tolerance for defining the 425.1368 peak was 10 ppm. The error bar represents one standard deviation of the triplicates run for this experiment. Background peaks appear at around 100 arbitrary units for the Fourier Transform mass spectrometer used here.

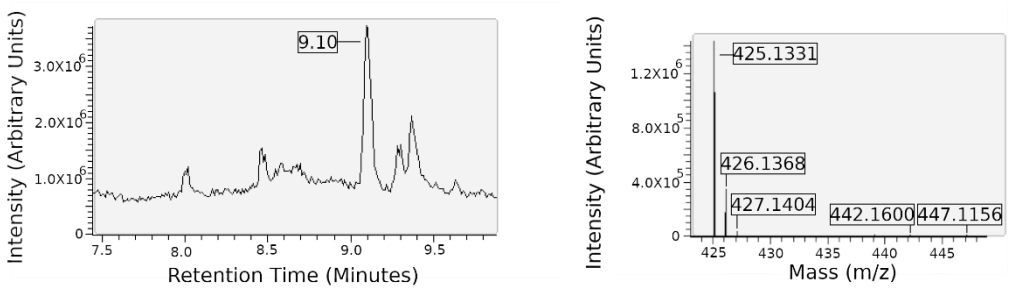


Figure 5. Mass peaks seen under the 9.10 retention time peak on the total ion chromatogram.

The 425.1368 mass compound seen in the *NRRL3_00042* overexpression strain produces peaks at 4 retention times: 8.13, 8.27, 9.10, and 10.26. The 9.10 retention time peak is the largest of the peaks by a factor of 89. All these peaks except the one at the 10.26 retention time disappear in the *NRRL3_00036* knockout. Average peak intensity at these retention times can be seen in figure 6.

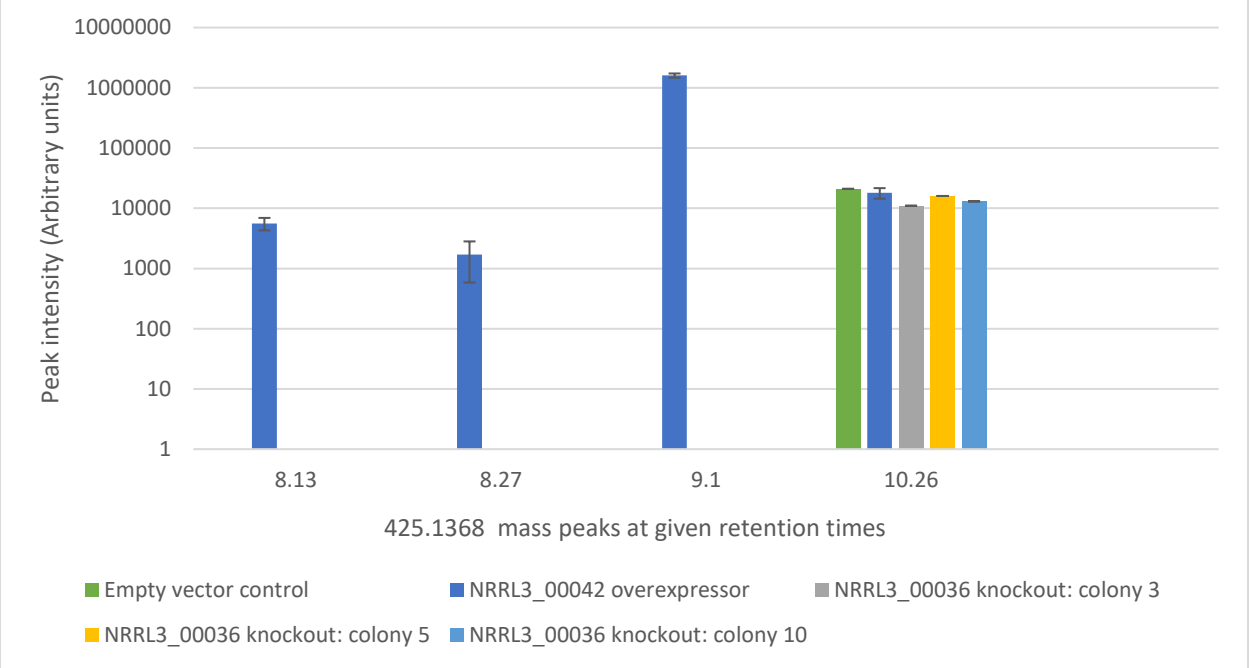


Figure 6. Average peak intensities of mass 425.1368 at different retention times across strains.

The 425.1368 mass appears at different retention times. These were included for the purposes of future identification, as the multiple retention times could mean that, like malformin, this compound could exist as different isomers. Units are reported as peak height in arbitrary units. Retention time is reported in minutes. The error bar represents one standard deviation of the triplicates run for this experiment. These results are from the same dataset used for figure 4.

The intensities of masses at the remaining retention times are quantified in figure 7. The chromatogram and mass peaks can be seen in figure 8.

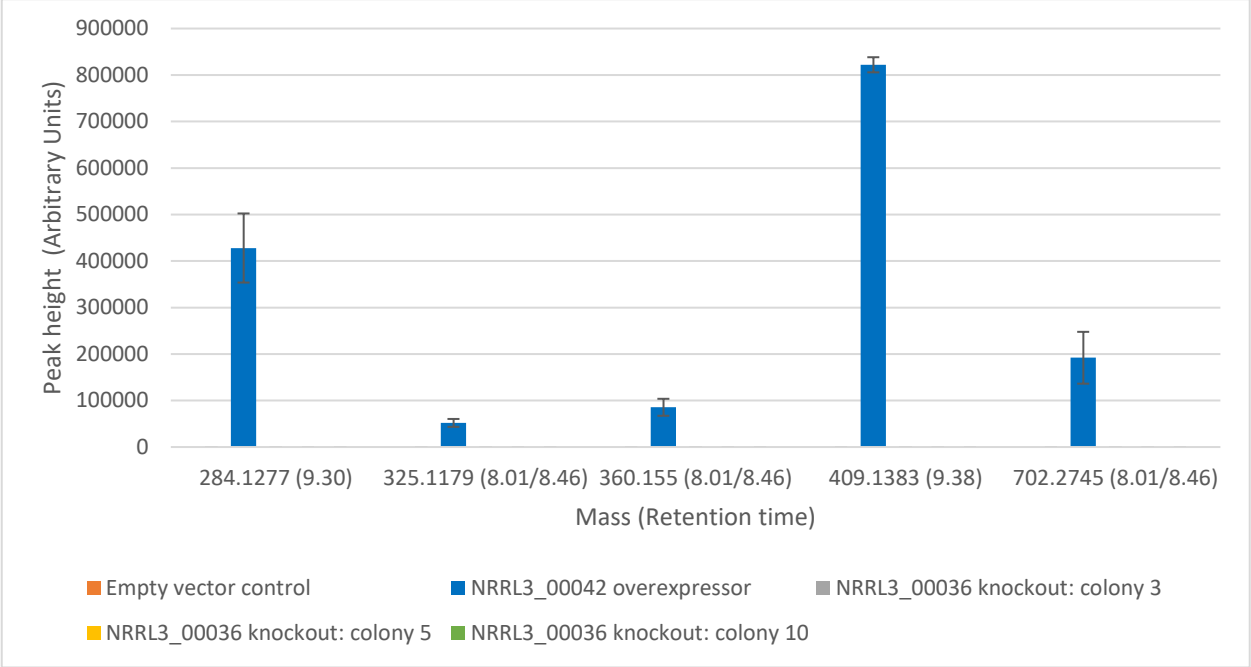


Figure 7. Other masses seen at labeled retention time peaks on the *NRRL3_00042* overexpression strain TIC.

Peak height for each retention time peak, of each mass to within 10 ppm, is shown in arbitrary units. The mass peaks shown here are were only observed in the *NRRL3_00042* overexpression strain. The error bar represents one standard deviation of the triplicates run for this experiment. These results are from the same dataset used for figure 4.

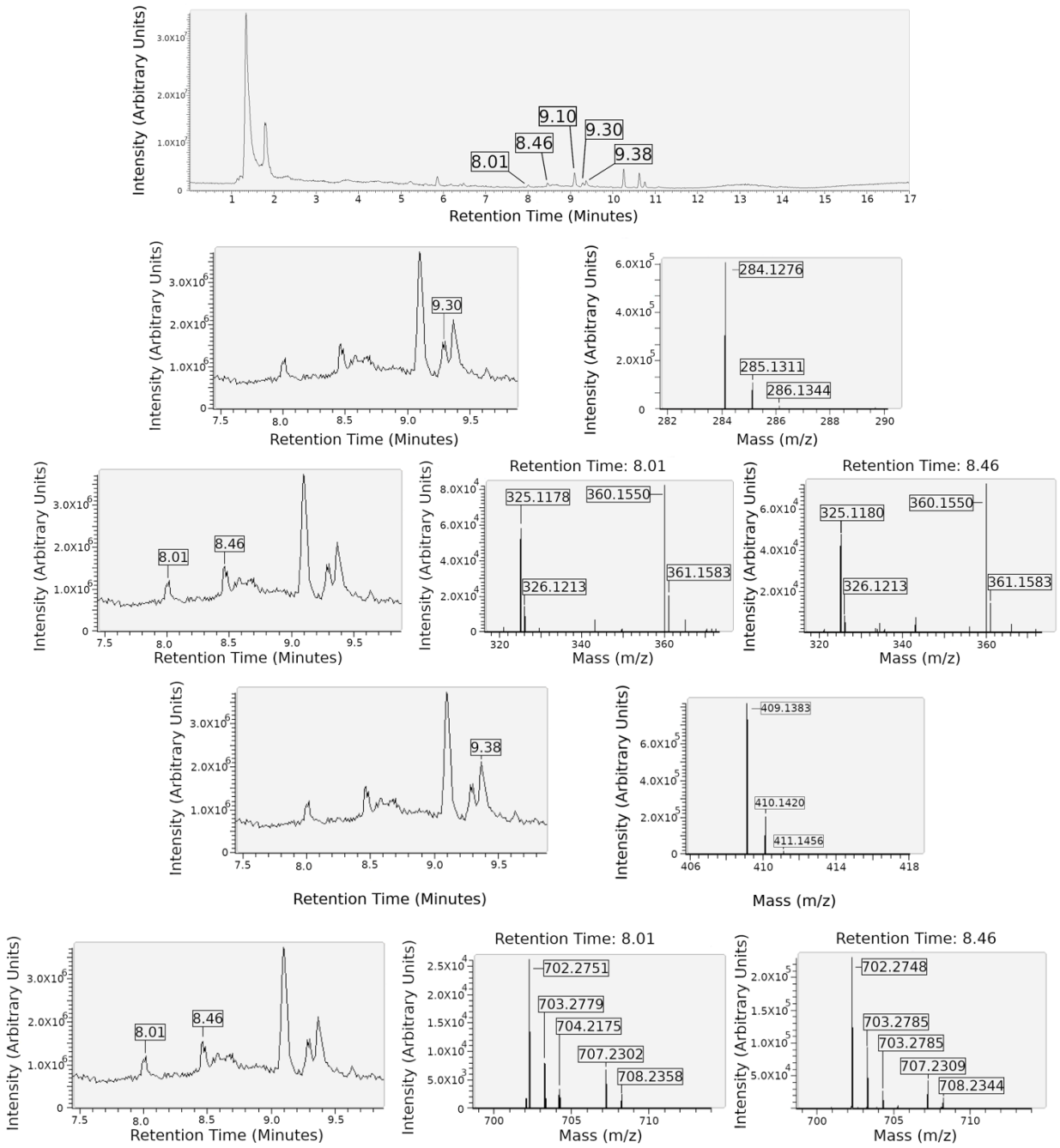


Figure 8. Unique mass peaks in the *NRRL3_00042* overexpression strain at the largest retention time peaks which disappear in the knockout.

Mass peaks are shown adjacent to the retention time peaks they were found under.

2.1.2 Phenotypes of the *NRRL3_00036* cluster strains

Phenotypic changes are evidence of underlying changes to an organism's genome. In the case of secondary metabolite producing fungi pigmentation could indicate production of a secondary metabolite.

Phenotypes of the *NRRL3_00036* cluster strains can be seen in figure 9. These phenotypes include changes in growth rate and pigmentation.

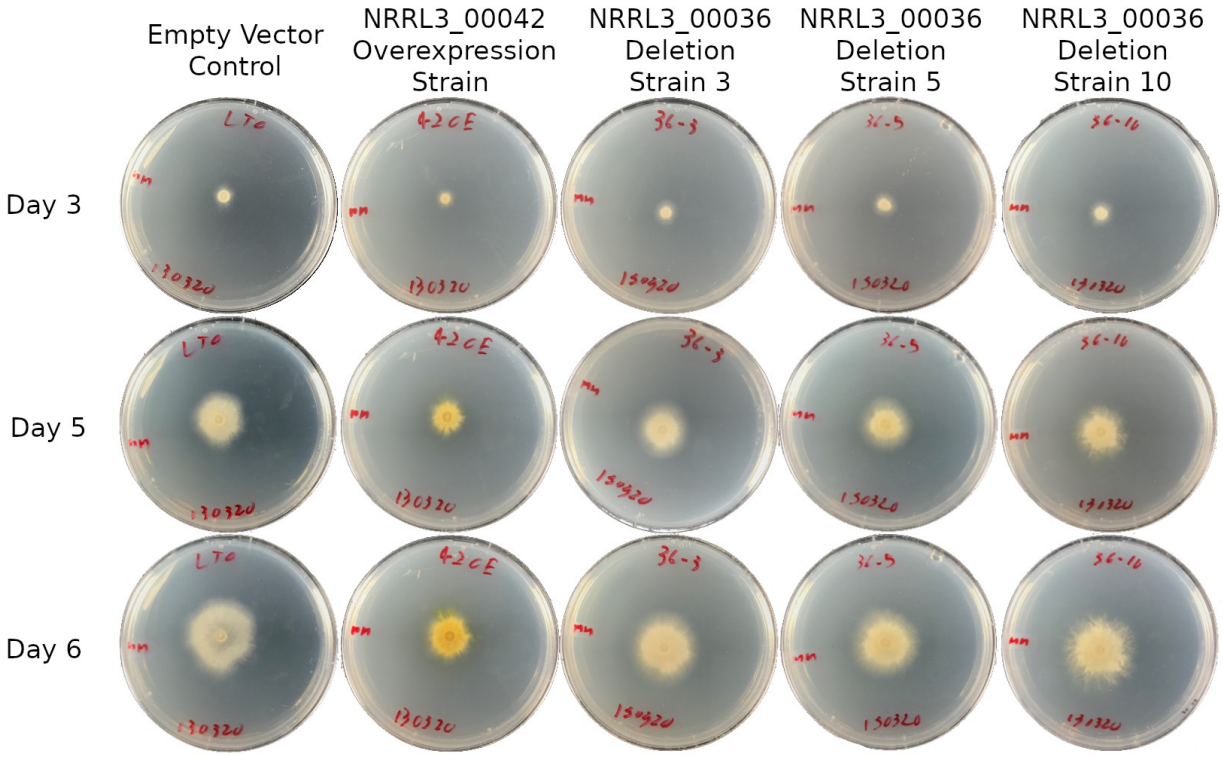


Figure 9. Time course of *A. niger* strains inoculated on minimal media + 1% maltose plates with $1X10^4$ spores grown at 30˚C.

2.1.3 *NRRL3_00036* cluster expression analysis

Reverse transcription PCR was run on the *NRRL3_00042* overexpression strain in order to demonstrate expression of the *NRRL3_00036* backbone. The gel from this reaction can be seen in supplemental figure 4. The gel shows a clearly noticeable band reverse transcribed and amplified from the *NRRL3_00036* gene. This band was not seen in the reverse transcriptase PCR of the parental strain.

In order to associate the appearance of metabolites to specific genes the expression of genes in the metabolite producing *NRRL3_00042* overexpression strain was examined (figure 10).

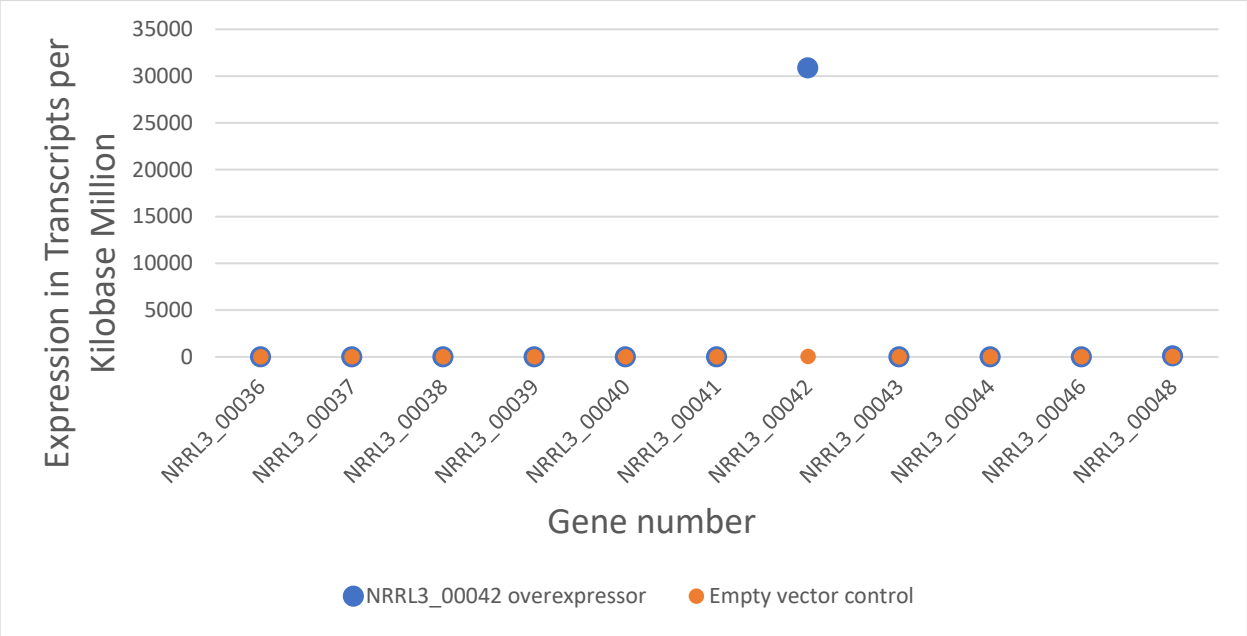


Figure 10. RNA expression values in TPM for the NRRL3_00036 cluster in the *NRRL3_00042* overexpression strain compared to empty vector control.

Only the gene cluster transcription factor *NRRL3_00042* showed upregulation. These results were obtained by extracting RNA from two-hour transfer cultures and running RNA-sequencing. The transfer culture is conducted by growing pre-cultures in 100 mL CM +2% fructose in 1 L flasks for 16 hours at 30°C and 220 rpm. Mycelia was collected by filtration with miracloth and washed with sterile ddH_2O. A half tablespoon of mycelia was transferred to 50 mL MM+1% cultures in 250 mL flasks. Cultures were grown for 2 hours at 30°C and 220 rpm and RNA, at which point was extracted for RNAseq.

2.1.4 *NRRL3_00036* transfer culture metabolite levels

In order to establish when metabolites production is highest in transfer culture, to determine the optimal time for RNA extraction, the *NRRL3_00042* overexpression strain was grown in a transfer culture in duplicate at several time points on MMJ. Each transfer culture sample was grown independently and discarded once that particular time point was collected. Results for the three identified novel masses can be seen in figure 11.

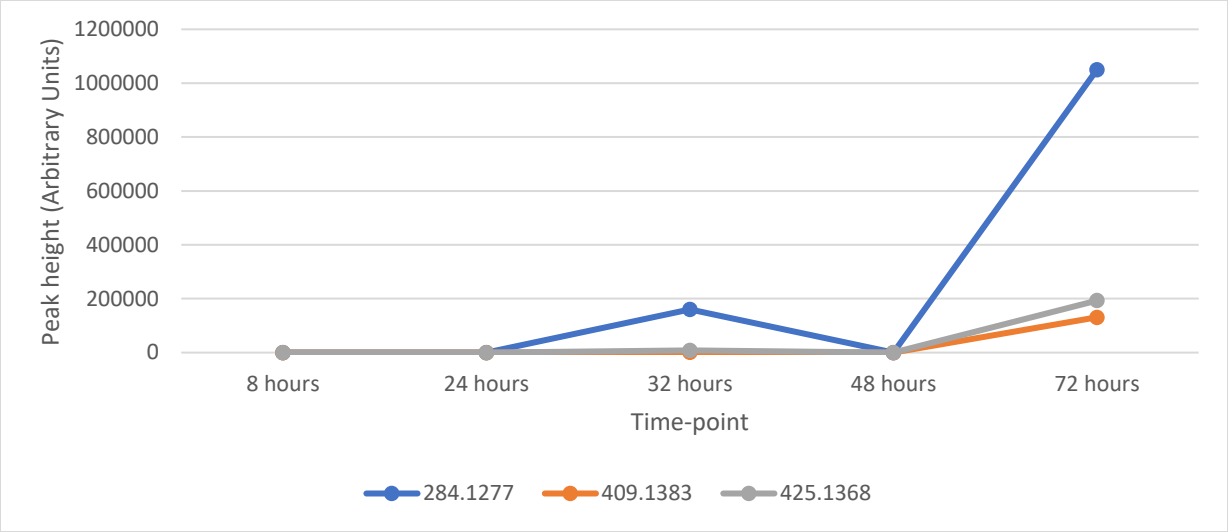


Figure 11. Time course of the average levels of novel masses in the *NRRL3_00042* overexpression strain in transfer cultures done in duplicate.

2.2 Malformins pathway and regulation

2.2.1 Putative malformins NRPSs

The production of malformins necessitates enzymes which produce them. For the purpose of its identification and characterization of its regulation a putative NRPS was selected for investigation.

A list of NRPS genes predicted from the NRRL 3 genome in was compiled as potential malformins producing enzymes (Table 5).

Table 5. List of NRPS genes present in the *A. niger* genome as determined bioinformatically by sequence homology.

Non-ribosomal peptide synthetase	Domain order
NRRL3_03167	A-T-C-Rpt-A-MT-A-T-C-A-T-Te
NRRL3_04180	A-T-C-A-T-C-A-T-C-A-T-C-A-T-C
NRRL3_05848	A-T-C-T-C
NRRL3_06801	A-C-T-MT-A-T-C-A-T-C-A-T-C-A-T-C
NRRL3_00036	T-C-A-T-C-Rpt
NRRL3_00135	A-T-C-C-A-T-C
NRRL3_01334	C-Rpt-A-T-C
NRRL3_07739	A-T-C-A-T-C
NRRL3_08341	A-T-C-A-T-NADB
NRRL3_08729	A-T-C-A-T-C
NRRL3_08790	A-T-C-A-T-C-A-T-C-A-T-C-A-T-C
NRRL3_08891	A-T-C-C-A-T-C-A-T-C-A-T-C-C
NRRL3_08969	A-T-C-A-T-C-A-T-C-C-A-T-C
NRRL3_08978	A-T-C
NRRL3_10148	A-T-T-C-Rpt-C-A-T-C-Rpt-(A)[MT}-A-T-C
NRRL3_10912	A-T-C-C-A-C-A-T-C-A-T-C-C-T-C-(T)
NRRL3_08538	A-T-C-A-T-C
NRRL3_11645	A-T-C-A-T-C-A-T-C-T-C-T-C

Legend :
A = adenylation, T = thiolation, C = condensation, MT = methyltransferase, Rpt = HxxPF repeat, NADB = NAD+ binding protein, Te = thioesterase/thioreductase

Domains were determined using an InterPro scans of the genes. The gene *NRRL3_08969* was selected as the best candidate NRPS for malformins synthesis.

2.2.2 *NRRL3_08969* cluster

Defining gene clusters allows us to look at the functions of those genes contained therein and relate them to the production of a metabolite or metabolites.

NRRL3_08969 belongs to a silent gene cluster in *A. niger* containing a four A domain iterative NRPS (*NRRL3_08969*) and a thioredoxin reductase (*NRRL3_8968*). This cluster was defined by starting from the NRPS backbone enzyme *NRRL3_08969* and working out towards nearby genes potentially involved with secondary metabolite clusters (59). This cluster can be seen in table 6.

Table 6. Secondary metabolite gene cluster in which the putative malformins backbone enzyme gene *NRRL3_08969* resides.

Gene Number	Gene Annotation
NRRL3_08964	aminotransferase class IV
NRRL3_08965.1	**fungal-specific transcription factor**
NRRL3_08966	MFS-type transporter
NRRL3_08967	MFS-type transporter
NRRL3_08968	Thioredoxin reductase
NRRL3_08969	**non-ribosomal peptide synthetase**
NRRL3_08970	S-adenosyl-L-methionine-dependent methyltransferase
NRRL3_08971	MFS-type transporter
NRRL3_08972	O-methyltransferase, COMT-type
NRRL3_08973	MFS-type transporter

Gene annotations have been manually curated (54).

2.2.3 Malformins Identification

In order to establish that malformins are being regulated and produced by certain gene clusters malformins must be correctly identified in the metabolome.

Malformin production was quantified by measuring mass spectrometry peaks from the metabolome of *A. niger* cultures. Identification of the peaks as malformins was done by looking for malformins adduct masses within 10 ppm, as well as the presence of the three distinct retention time peaks seen in publications on malformins, such as in the Theobald *et al*. paper (10). A comparison of total ion chromatograms (TICs) from these strains can be seen in figure 12. The extracted ion chromatograms at the malformins masses can be seen in figure 13.

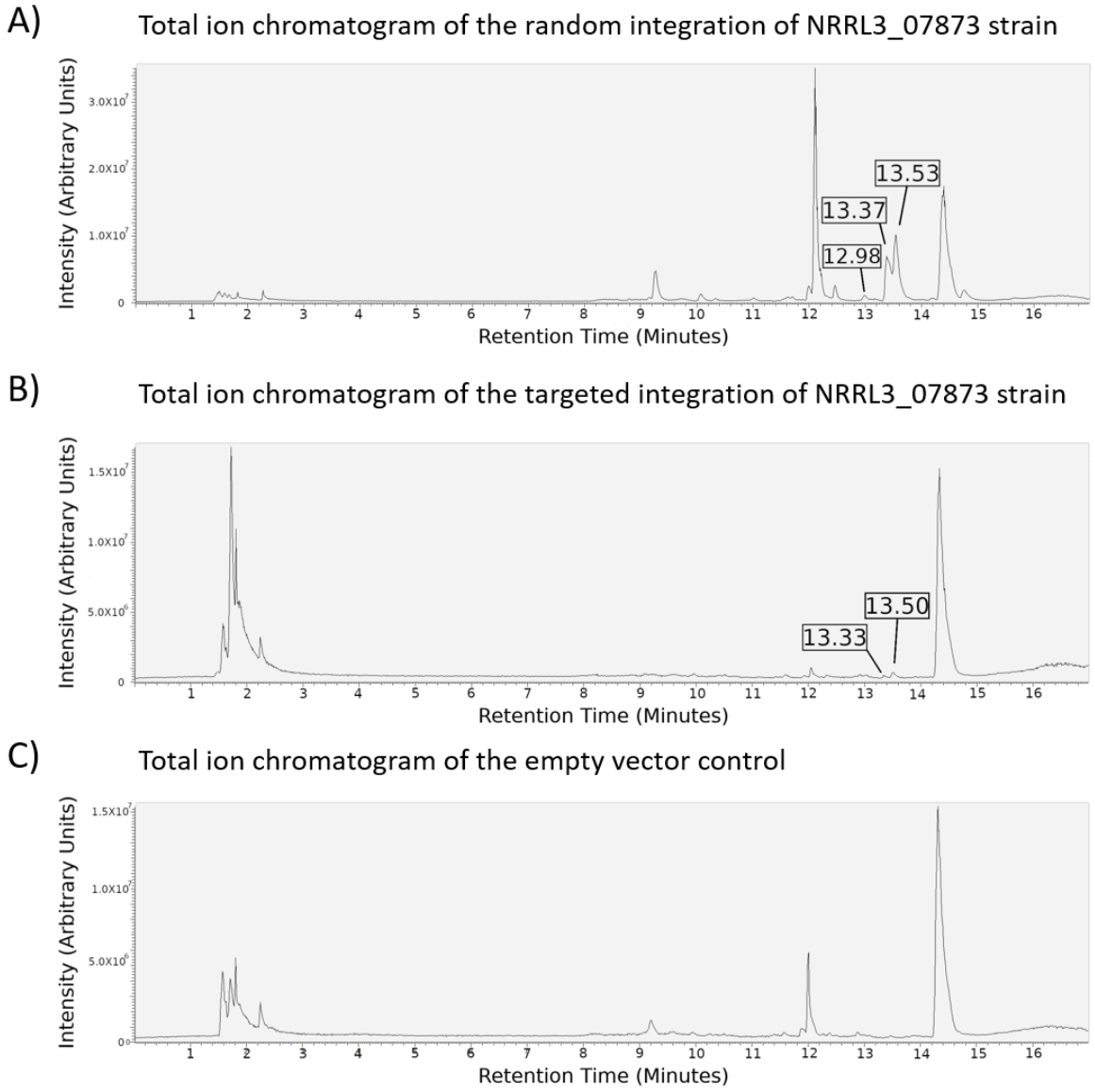


Figure 12. Flight time mass spectrometry results of *A. niger* metabolome from cultures grown 5 days, 30°C, on MMJ in 96 well plates. A) Total ion chromatogram of *NRRL3_07873* random integration strain. B) Total ion chromatogram of *NRRL3_07873* targeted integration into GlaA locus strain. C) Total ion chromatogram of the empty vector negative control.

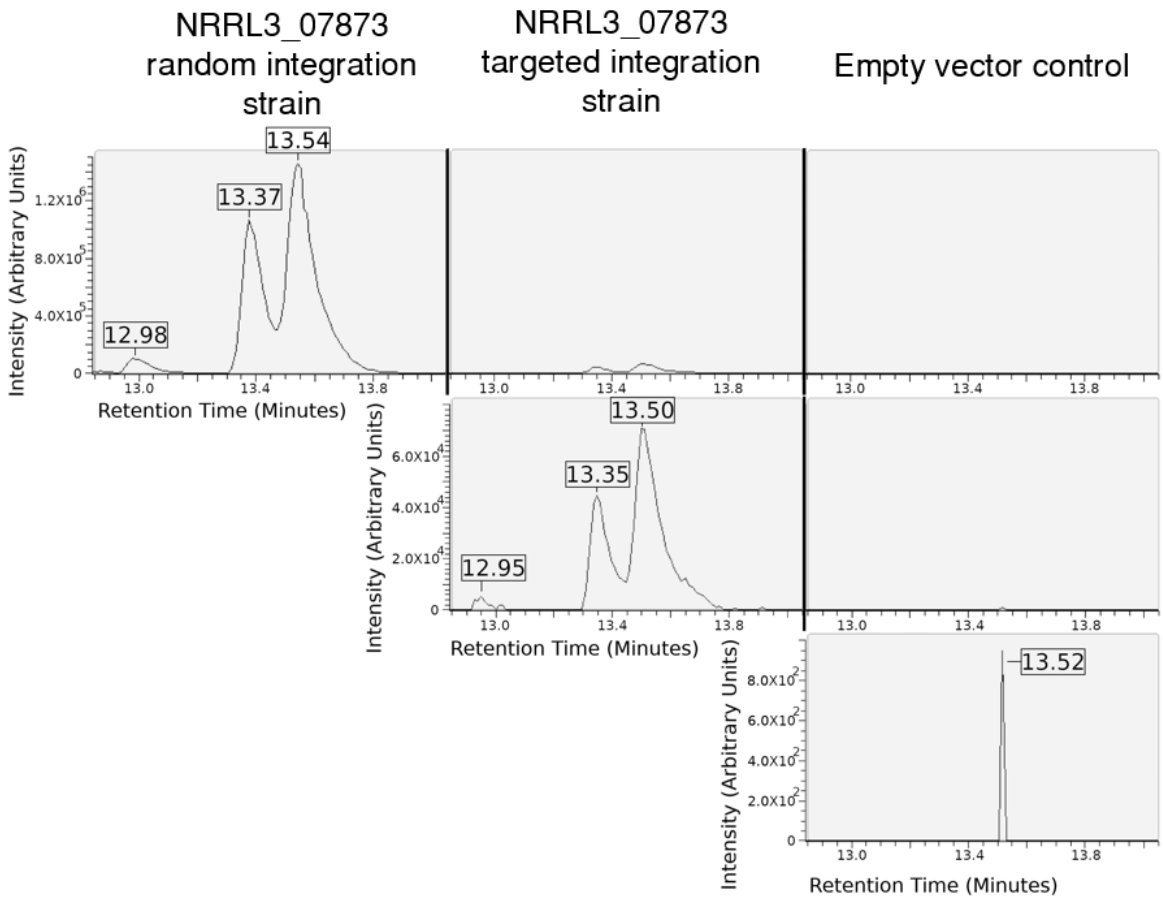


Figure 13. Extracted ion chromatograms of the masses 530.2469 and 516.6959 with a 10 ppm tolerance at a retention time window of 12.84-14.05 at three different scales.

The empty vector control from figure 13 and was taken from 1 of the 3 negative controls that had a peak at the 13.5 retention time for the 530.2469 malformins mass, out of the twelve negative controls run. This is compared to 6 of the 9 *NRRL3_08965.1* targeted integration strain samples having the 13.5 retention time peak and 9 of 9 *NRRL3_07873* targeted integration strain samples having the peak.

The individual mass peaks at the malformins retention times can be seen in figure 14. Out of the three negative controls which a contained 13.5 retention time peak for the 530.2469 malformins mass one also showed the three malformins adducts shown in figure 14A, though without the additional isotopic masses. The other two only contained a single mass peak within 10 ppm of 530.2469, as seen in figure 14C.

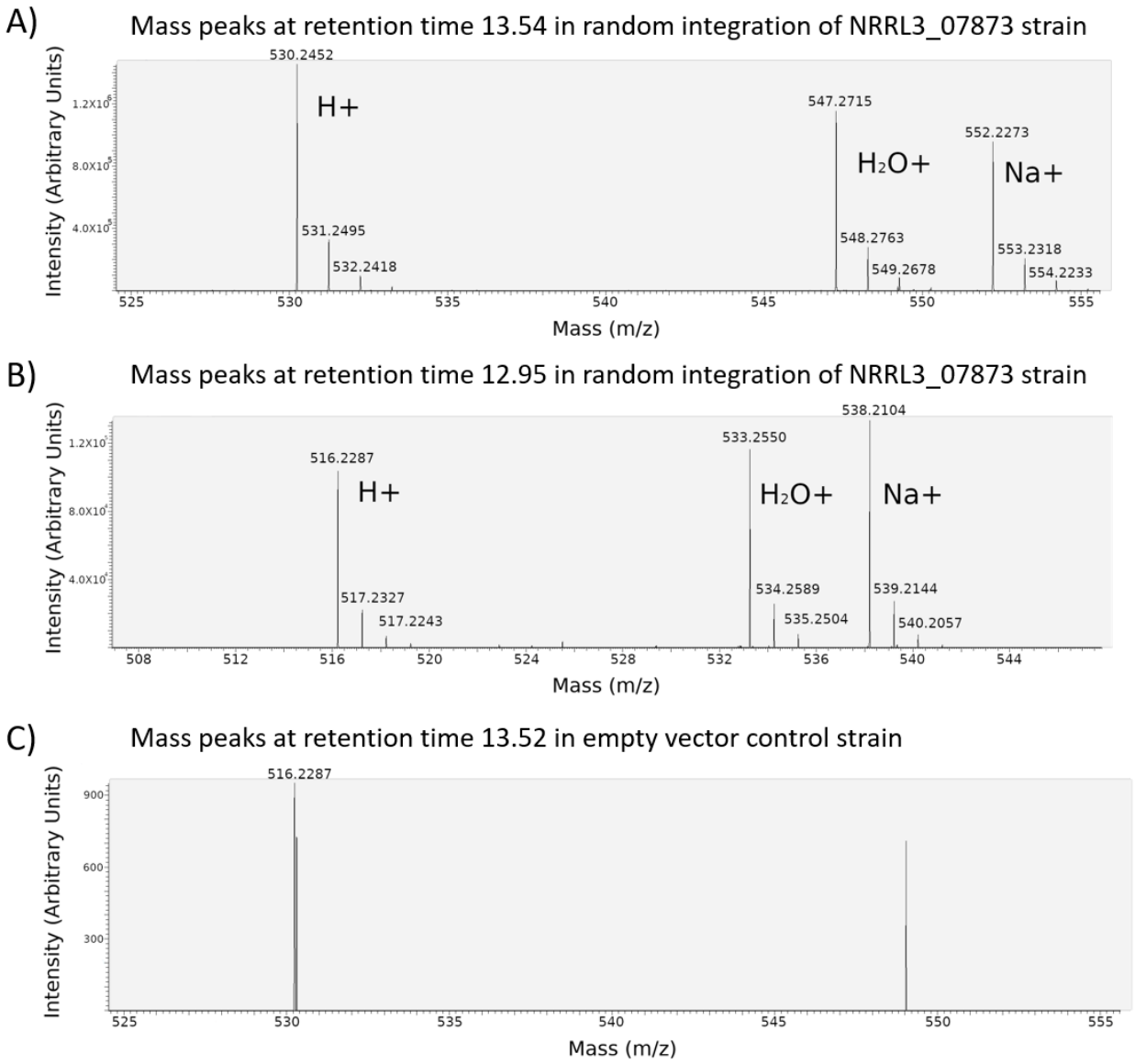


Figure 14. Mass peaks at given retention times with malformins adducts labelled. A) Mass peaks at retention time 13.54 from the TIC in figure 12A. B) Mass peaks at retention time 12.95 from the TIC in figure 12A. C) Mass peaks at retention time 13.52 from the TIC in figure 12C.

2.2.4 Malformins levels

Upregulation of gene cluster transcription factors are expected to increase the expression of the cluster genes in which they reside. Since the *NRRL3_08969* cluster contains genes thought to be involved in malformin production the levels of malformin in the *NRRL3_08965.1* overexpression strain was investigated. Since other transcription factors were upregulated (as part of a larger, unpublished, project) malformins levels were investigated in these strains as well. The *NRRL3_07873* overexpression strains are included in these results because they gave the highest malformins levels.

Initial results for malformins production were obtained from random integration strains where a number of different transcription factors with a *GlaA* promoter were randomly integrated into

the *A. niger* genome along with a *pyrG* selective marker. From these random integrations, 11 strains containing the *NRRL3_07873* transcription factor were generated. Taking one sample from each of the 11 strains gave an average malformins peak intensity of $9.2X10^5$ with a min of $1.6X10^3$, a max of $2.5X10^6$, and a standard deviation of $9.0X10^5$. Twelve other transcription factors were also randomly integrated to make new strains. These other strains showed lower levels, some even lower than the negative control. A new strain was generated with the same random integration vector containing NRRL3_07873, which resulted in an average peak intensity of $1.0X10^6$, a minimum of $4.4X10^5$, a maximum of $1.5X10^6$, and a standard deviation of $3.2X10^5$. The malformins produced in various samples and strains can be seen in figure 15. A list of malformins production values in different strains can be seen in table 7.

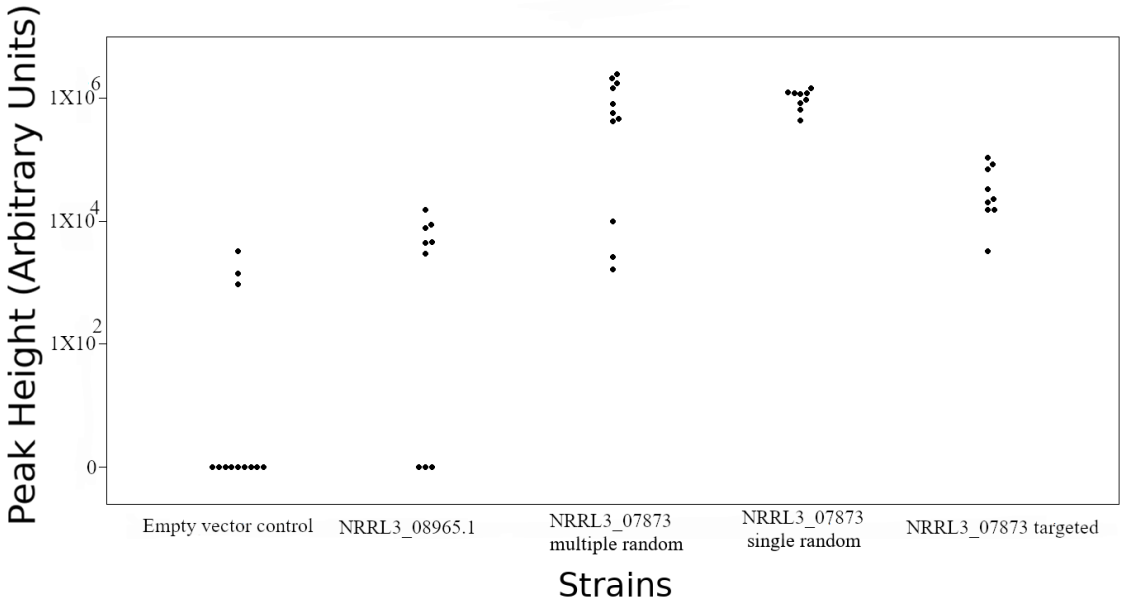


Figure 15. Beeswarm plot of malformins peak height in arbitrary units according to strain.

Included in these results are the strains where transcription factors were inserted into the glucoamylase locus ("NRRL3_08965.1", and "NRRL3_07873 targeted") as well as a set of eleven independent *NRRL3_07873* random integrations ("NRRL3_07873 multiple random") and the random integration strain generated for this project ("NRRL3_07873 single random"). The Y-axis has a base 10 logarithmic scale representing peak height in arbitrary units and each point represents an individual sample.

Table 7. Malformins production statistics across notable strains.

Strain	Min	Max	Average	P-value (two-tail t-test assuming equal variance compared to empty vector)
Empty Vector control	0	$3.30X10^3$	$6.3X10^2$	1
NRRL3_08965.1	0	$1.56X10^4$	$4.92X10^3$	0.020

NRRL3_07873 - multiple independent random integration strains	$1.6X10^3$	$2.5X10^6$	$9.2X10^5$	0.0066
NRRL3_07873 - single random integration strain	$4.44\ X10^5$	$1.45X10^6$	$1.02X10^6$	$6.17X10^{-8}$
NRRL3_07873 integration into GlaA locus	$3.23X10^3$	$1.08X10^5$	$4.14X10^4$	0.0042

Table 7 P-values come from three sets of triplicates for the *NRRL3_08965.1* targeted integration strain, *NRRL3_07873* targeted integration strain, and *NRRL3_07873* single random integration strain compared to three triplicates of the negative control. The pooled data from multiple *NRRL3_07873* random integration strains used 11 strains each grown once compared to three triplicates of the negative control. Values represent mass spectrometry peak heights and are in arbitrary units.

2.2.5 *NRRL3_07881*/*NRRL3_07884* Cluster

The functional role of the *NRRL3_07881*/*NRRL3_07884* cluster genes was examined for involvement in the malformins pathway by examining the annotation of cluster genes.

The transcription factor gene *NRRL3_07873* belongs to the silent *NRRL3_07881*/*NRRL3_07884* gene cluster. This cluster can be seen in table 8.

Table 8. Secondary metabolite gene cluster in which the transcription factor gene *NRRL3_07873* resides.

Gene Number	Gene Annotation
NRRL3_07873	**fungal-specific transcription factor**
NRRL3_07874	zinc-type alcohol dehydrogenase
NRRL3_07875	acyl-CoA N-acyltransferase
NRRL3_07876	hypothetical protein
NRRL3_07877	FAD-binding domain-containing protein
NRRL3_07878	tannase/feruloyl esterase family protein
NRRL3_07879	ankyrin repeat domain-containing protein
NRRL3_07880	isochorismatase family protein
NRRL3_07881	non-ribosomal peptide synthetase-like protein
NRRL3_07882	phenylalanine ammonia-lyase
NRRL3_07883	thioesterase domain-containing protein
NRRL3_07884	polyketide synthase
NRRL3_07885	AMP-dependent synthetase/ligase

Gene annotations have been manually curated (54).

2.2.6 External stimuli and environmental conditions for malformins production

Given that the malformin production observed in the random integration strain was not being reached in the *NRRL3_07873* targeted integration strain when both have the same parental strain the involvement of other regulatory pathways was investigated. In order to further characterize malformins regulation a number of different inducing conditions were tested on the laboratory NRRL 2270 parental strain that was used to generate the malformins producing strains. All samples were done in triplicate. No conditions showed induction of malformins production at levels deemed worth investigating. The list of tested conditions can be seen in table 9.

Table 9. Conditions tested on the parental NRRL 2270 strain in an attempt to induce malformins production.

Media	Light	NaCl	Temperature	Duration
MM+ 1% maltose	Grown in 24 well plate in incubator	0%	30°C	7 day
MM+ 1% maltose+ 1%Wheat	Grown in 24 well plate in incubator wrapped in aluminum foil	5%	37°C	10 day
Yeast Malt Extract	Grown in in 24 well plate on a hot plate by the window			

The effect of 5% salt concentration and agar media on several malformins producing strains was tested. Two separate *NRRL3_07873* random integration strains as well as the *GlaA* targeted *NRRL3_07873* integration strain and the empty vector Anep8 negative control strain were tested. The "2" random integration is one generated for this study while the "S" strain was produced prior to this study. All samples were done in triplicate. The malformins production values for these strains can be seen in table 10.

Table 10. Conditions tested across strains in an effort to increase malformins production.

Media	Strain	NaCl	Temperature	Time	Malformins Peak Height
MM+1% maltose agar	NRRL3_07873 random integration -2	0%	30°C	8 day	$5.30X10^5$
MM+1% maltose agar	NRRL3_07873 random integration -S	0%	30°C	8 day	$3.88\ X10^5$
MM+1% maltose agar	NRRL3_07873 targeted integration	0%	30°C	8 day	0

MM+1% maltose agar	Empty vector control	0%	30°C	8 day	0
MM+1% maltose agar	NRRL3_07873 random integration -2	5%	30°C	8 day	3.12 $X10^5$
MM+1% maltose agar	NRRL3_07873 random integration -S	5%	30°C	8 day	3.25 $X10^5$
MM+1% maltose agar	NRRL3_07873 targeted integration	5%	30°C	8 day	0
MM+1% maltose agar	Empty vector control	5%	30°C	8 day	0

Cultures were grown in 24 well plates on 2.5 mL of agar minimal media +1% maltose. Agar was grinded in an equal volume of methanol to extract metabolites. Peak height is given in arbitrary units.

2.2.7 Expression analysis

Since malformins expression necessitates the expression of genes the expression profiles of the novel targeted integration strains were investigated. Specifically, NRPSs, including the putative malformins NRPS, were investigated for upregulation.

RNA-sequencing was used to quantify RNA expression in the strain where *NRRL3_07873* was integrated into the glucoamylase locus. RNA was collected after a two-hour transfer culture in *GlaA* promoter inducing medium containing maltose. Results were reported in transcripts per kilobase million (TPM). This showed upregulation of the *NRRL3_07873* transcription factor but not its cluster or the putative malformins cluster. This was also done with *NRRL3_08965.1* overexpression strain and an empty vector control. Neither showed upregulation of the *NRRL3_08969* NRPS gene or its cluster. Results for the *NRRL3_07881*/*NRRL3_07884* cluster can be seen in figure 16, while those for the *NRRL3_08969* cluster can be seen in figure 17.

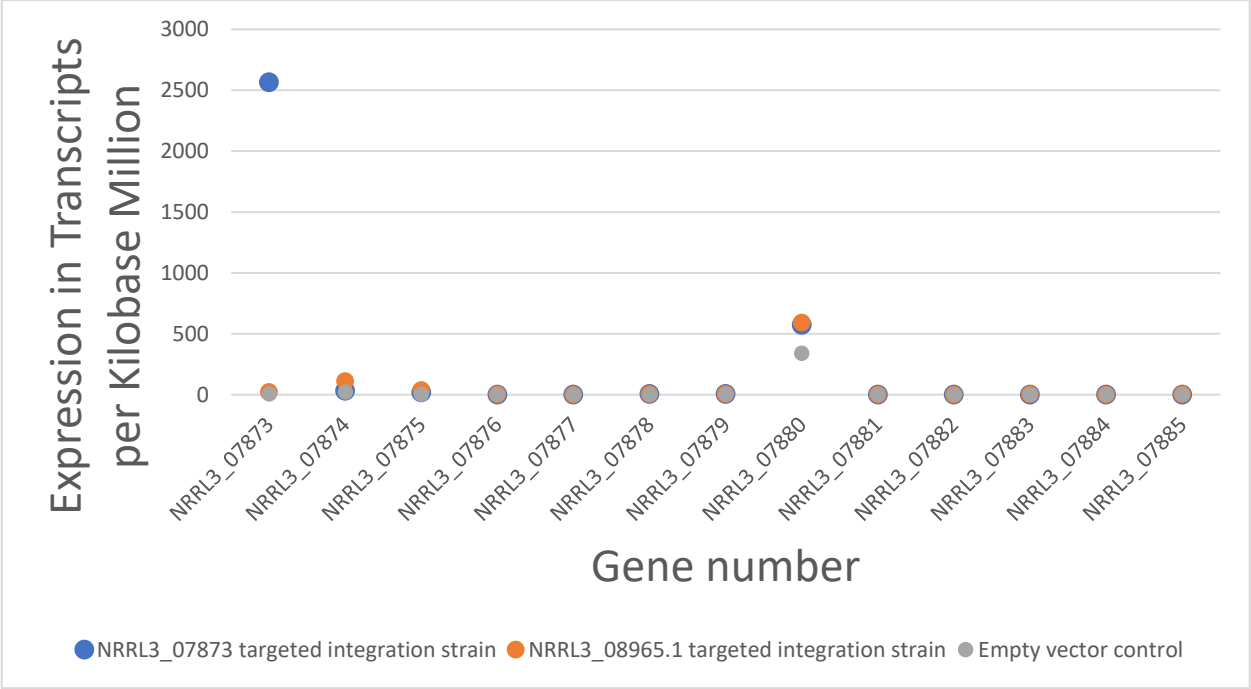


Figure 16. RNA expression values in TPM for the *NRRL3_07881*/*NRRL3_07884* cluster.

Only the gene cluster transcription factor *NRRL3_07873* showed upregulation, besides *NRRL3_07880* which showed a less than 2-fold upregulation in both overexpression strains. These results were obtained by extracting RNA from two-hour transfer cultures and running RNA-sequencing.

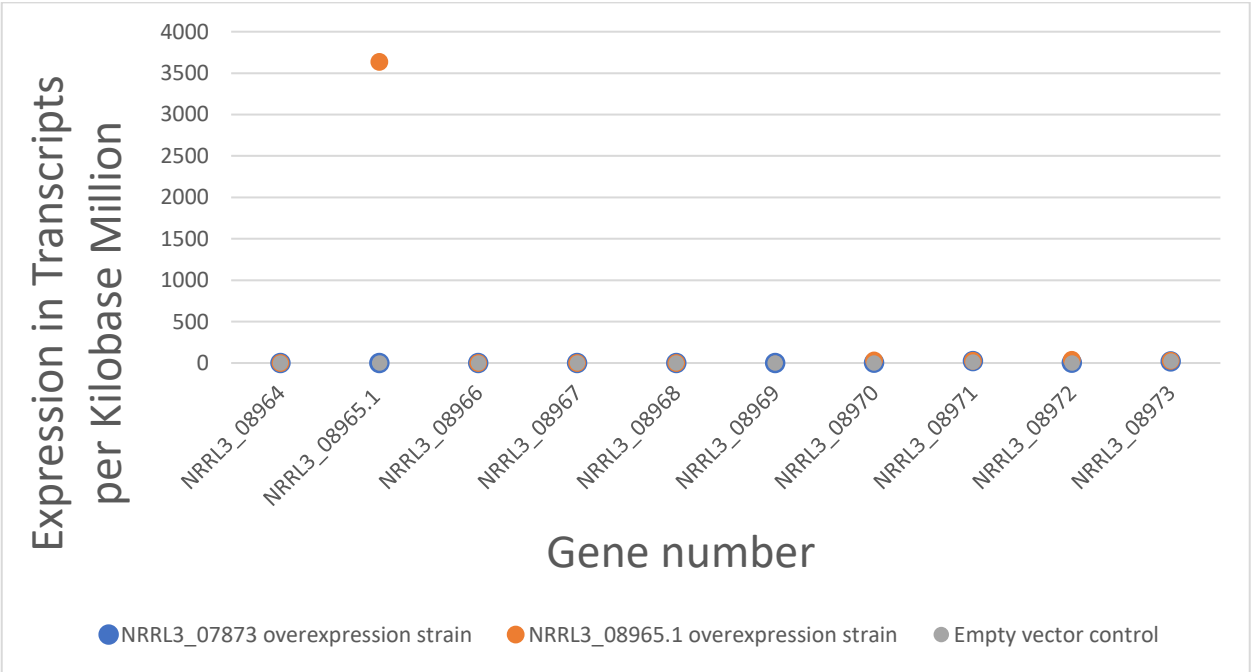


Figure 17. RNA expression values in TPM for the *NRRL3_08969* cluster.

Only the gene cluster transcription factor *NRRL3_08965.1* showed upregulation. These results were obtained by extracting RNA from two-hour transfer cultures and running RNA-sequencing.

2.2.8 Transfer culture malformins production

If there is a time frame between induction and gene expression, then any given time point taken for RNA expression may not show expression of the relevant genes. In order to find a better time point for RNA expression a time course transfer culture, run under the same conditions as done for RNA sequencing, was conducted.

The malformins production, the anticipated consequence of this RNA expression, of two strains at different transfer culture time points was measured in order to determine how much of a delay there is between induction and malformins expression under these culture conditions (figure 18). This showed peak malformins expression at 24 hours.

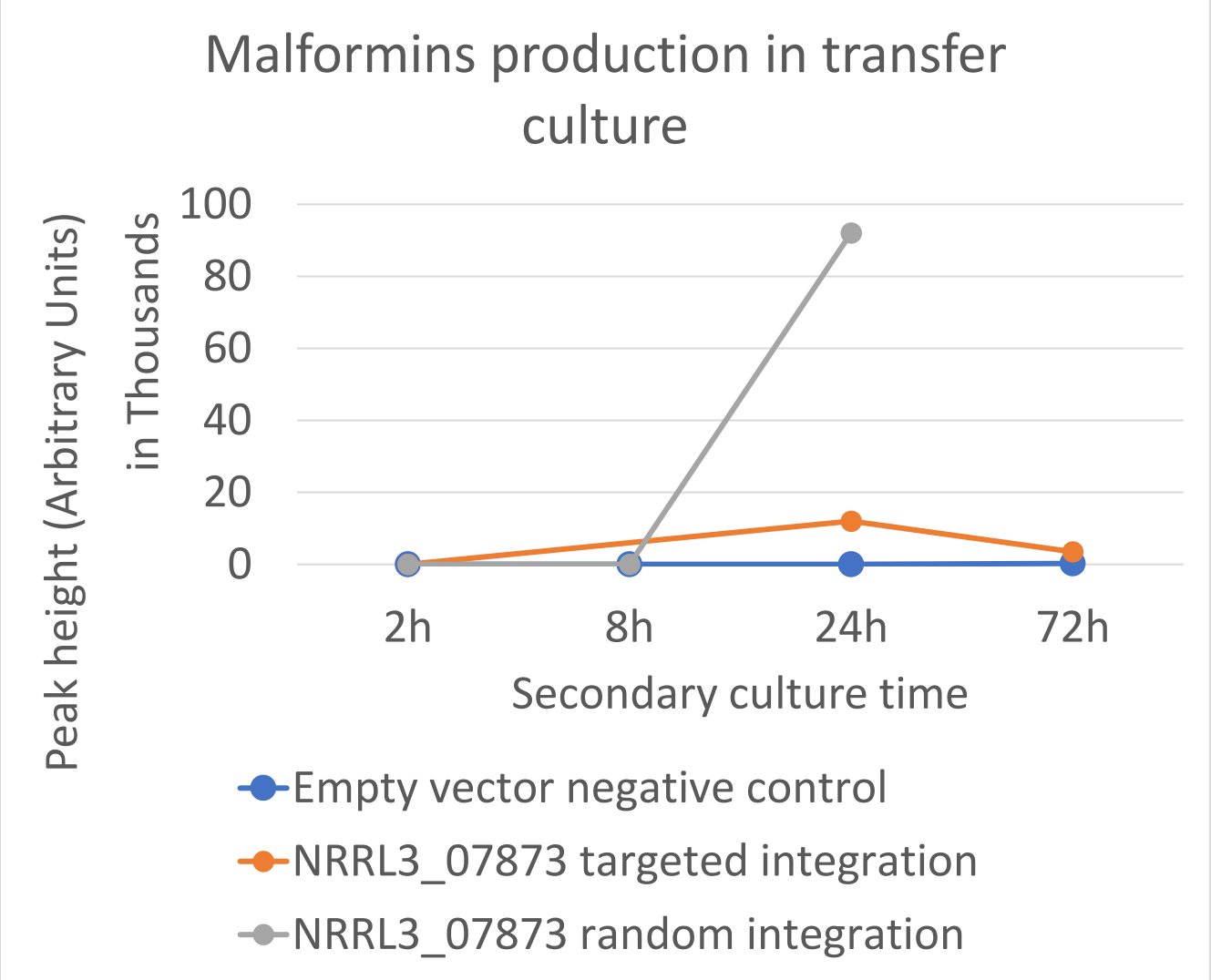


Figure 18. Time course of malformins levels in transfer culture.

2.3 McrA depletion strain metabolomics

Given that the targeted integration strain does not produce as much malformins as NRRL 2270 derived strains are capable of producing, as demonstrated by the random integration strains, some other regulatory mechanisms may be involved. If any single gene is to be selected for testing the greatest chance of observing a positive effect would be by investigating genes which regulate many gene clusters. To that end an *McrA* depletion strain was generated.

Principle component analysis was conducted on two triplicates of the empty vector control strain and a strain where 175 base pairs of the beginning of the gene and 310 base pairs of the promoter upstream of the *McrA* gene were removed (figure 19). Some masses observed uniquely in the *McrA* depletion strain, or masses highly upregulated in said strain, that were present at high levels, are listed in table 11.

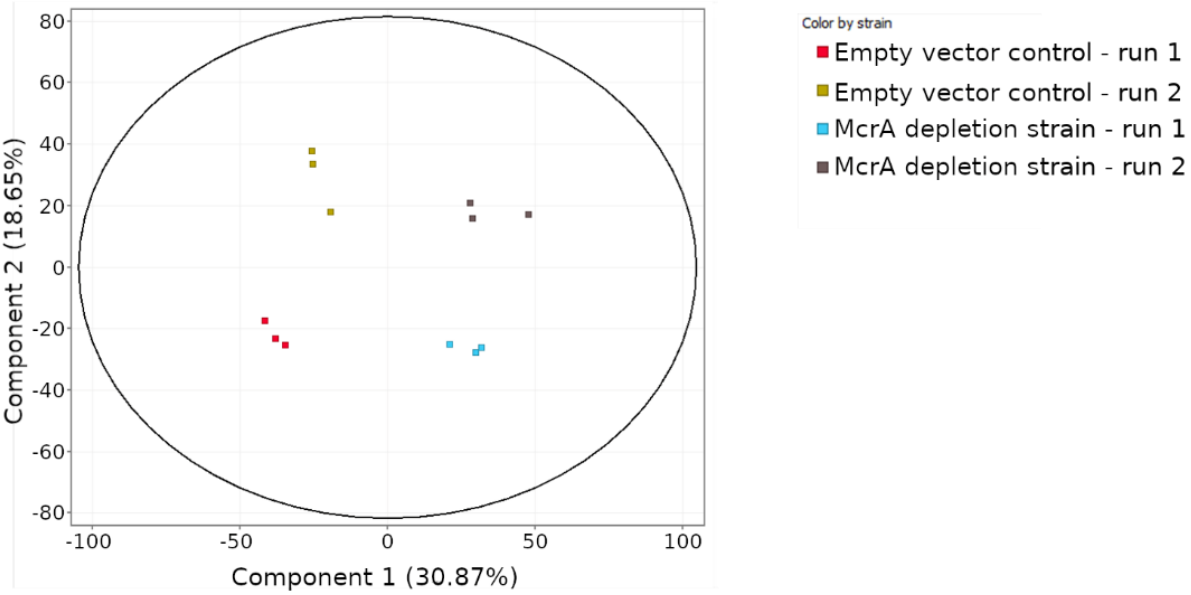


Figure 19. Principle component analysis of two runs of the McrA depletion strain against the empty vector control.

Strains were grown on 190 µL of MMJ in 96 well plates for 5 days at 30˚C. Each run was done in triplicate, with each sample visible as an individual point on the principle component analysis plot. Proportion of the variance is displayed as a percentage nest to the component number. Entities were detected by a quantitative time of flight mass spectrometer. Background peaks typically appear at 100 arbitrary units, but entities below 1000 arbitrary units were omitted from analysis. Mass Profiler Professional software was used to generate the PCA.

Table 11. Compounds most upregulated in the *McrA* depletion strain relative to the empty vector control, with peak intensity measured in arbitrary units

Compound with equivalent mass	Mass	Type	Empty vector control peak average	*McrA*(-) peak Average	Retention time
Asticolorin A	578.19	Mycotoxin not observed in *A. niger*	0	$1.2X10^5$	1.27
Mannopine	310.14	Opine observed in Plant/Bacteria interactions	2883	$1.2X10^5$	4.53
porphobilinogen	208.09	Intermediate of porphyrin biosynthesis	0	$1.6X10^5$	1.69
Piperaduncin B	510.163	Anti-bacterial Flavanoid	0	$6.7X10^4$	1.27
Miraxanthin-I	181.028	Yellow plant dye	0	$5.3X10^4$	2.41
Diflunisal	136.013	Anti-inflammatory	8,645	$4.7X10^4$	3.39

Some highly upregulated compounds were selected to further characterize the novel strain. The compounds named are compounds which share a mass with the listed mass peaks and were chosen by IDBrowser software that comes with Mass Profiler Professional using an internal metabolite database. Peak values are displayed as peak height in arbitrary units.

Discussion

3.1 Putative NRRL3_00036 backbone enzyme metabolite pathway analysis

Given that the *NRRL3_00042* overexpression strain produces a number of compounds at higher levels than the empty vector control, knockouts on this strain can be performed to determine which genes are involved. Of particular interest is the backbone enzyme situated in the same cluster, *NRRL3_00036*. Knock-out of this gene is expected to eliminate both the yellow color the strain produces and at least one of the unique compounds seen by mass spectrometry.

NRRL3_00036 is predicted to be an NRPS, and so is expected to produce an NRP (70). NRPs are made of colorless amino acids (71), so the yellow pigment seen in the *NRRL3_00042* overexpression strain is possibly the result of a modification from a tailoring enzyme. It could also be a modification performed by a domain on the NRPS enzyme itself, such as the Ox domain seen in the indigoidine synthesizing NRPS (72). The NRRL3_00036 shows no optional domains by InterPro scan. This suggests more enzymes than just the backbone enzyme in the pathway. While outside the scope of these experiments, this gives further research avenues into deciphering this particular pathway.

Knockout of the *NRRL3_00036* gene has not yet been confirmed by spanning primers. Flanking primers show no amplicons (supplemental figure 3), even though the same primers show the expected length amplicon in the *NRRL3_00042* overexpression strain. A set of spanning primers failed to amplify the deleted region, which would suggest either a larger deletion than expected, or an issue with that particular set of primers. Even without a larger than expected deletion these deletion strains are expected to remove the last 6 base pairs of the upstream hypothetical protein gene (*NRRL3_00035*) whose stop codon is 348 base pairs upstream of the *NRRL3_00036* start codon. The downstream *NRRL3_00037* gene is 721 base pairs away and so is not expected to be affected unless some kind of nearby regulatory element was disrupted. Metabolomic and phenotypic evidence indicates that some kind of genetic alteration occurred.

Removing the *NRRL3_00036* gene eliminates most of the yellow color seen in the *NRRL3_00042* overexpression strain but still leaves the mycelia more yellow than the empty vector control.

Total ion chromatograms show five visible retention time peaks that are not present in the parental strain and abolished in the *NRRL3_00036* knockout. Three of the masses were not identified by a search of an internal metabolome database, or from FoodDB. The 284.1277 mass did show several compound matches within 10 ppm on FoodDB, none of which were identifiable as fungal metabolites. The other three masses, 325, 360, and 702, are present at the same two retention times. This suggests that there might be some structural similarity. Mass differences were compared to adduct masses to determine if they were the same compound. The 702 mass is nearby a 708 mass peak, which would suggest that they are respectively NH_4+ and Na+ adducts, which would suggest a compound mass of 684. If we consider that this might be the same compound ionized to itself, then the real mass of the compound would be 342. This is an NH_4+ mass away from 360 while the mass at 325 could be the loss of a water molecule and the gain of a proton. This gives a good indication that these are all the same compound, and one with the same mass as maltose, which is not unexpected given that the fungi were grown on maltose as a carbon source. That only the overexpression strain contained maltose when all strains were grown for the same length of time with the same starting spore count with spores of the same age indicates that the overexpression strain consumes maltose less quickly. This, in addition to carbon being diverted towards the other unique compounds, helps explain why the *NRRL3_00042* overexpression strain grows less quickly. The precipitation step in sample preparation for mass spectrometry can cause

large variations in maltose levels, so a better designed experiment would need to be conducted to corroborate this observation. The unidentified three masses (284, 409, and 425) do not appear to be adducts of each other.

An additional line of evidence pointing to the involvement of *NRRL3_00036* in the production of these peaks is its increased expression in the *NRRL3_00042* overexpression strain, as seen in by reverse transcription PCR in supplemental figure 4. Here, the *NRRL3_00036* band is more intense in the *NRRL3_00042* overexpression strain compared to the parental strain when grown for the same amount of time and amplified for the same number of cycles. Also seen is greater expression at 7 days of growth than 5 days of growth. Reverse transcriptase PCR is sensitive, able to detect as little as 100 copies of RNA (73), so even basal expression of the backbone enzyme could be picked up given enough cycles of amplification. This could help explain why the identified novel masses do not appear on mass spectrometry in the parental strain. Supplemental figure 5 is also included to show relative levels of expression of the *NRRL3_00036* and *NRRL3_00042* genes. Real-time quantitative reverse transcription PCR could be used to give more accurate results

3.1.1 *NRRL3_00042* overexpression strain expression analysis and transfer culture

The lack of expression of cluster genes for the *NRRL3_00042* overexpression strain may be because the two-hour time point for RNA extraction after induction by maltose was too soon for gene cluster expression to occur. In order to determine when the production of the masses of interest occurs in the transfer culture a transfer culture time course was conducted. The data available does not point to a clear trend in production but does show that the masses only appear starting at the 36-hour time point. This is well after the two hours used for expression analysis and so it is hypothesized that taking RNA at a later time point could show upregulation of the cluster.

3.2 Malformins pathway and regulation

3.2.1 Candidate NRPS gene for malformins pathway

NRPs are produced by NRPSs (70). If malformins is an NRP then it would follow that an NRPS is required for its production. Alternatively, malformins could be ribosome-dependently supplied peptides modified by epimerases to convert L amino acids to D amino acids (74). Given that malformins contains non-proteinogenic amino acids and exists in multiple, closely related, forms, it is more likely a non-ribosomal rather than a ribosomal peptide, since translation is a more accurate process than NRP synthesis. In eukaryotes, the rate of misincorporation of amino acids varies from 10^{-3} to 10^{-6} (75), a rate too low to produce the 9 forms of malformins, given that two of the retention time peaks are approximately equivalent in size (as seen from the 13.35 and 13.50 minute retention time malformins peaks). This, combined with existing literature showing evidence a homologous NRPS produces malformins in *A.brasiliensis* (10) , makes pursuing a putative NRPS the strategy with the greatest chance of success.

As *A. niger* is a known malformins producer there is a good chance that one of the 18 known non-ribosomal peptide synthetases in the *A. niger* genome is responsible for its production. While there are also NRPS-like enzymes present in the genome, these enzymes do not produce peptides. This is because: 1) NRPS-like enzymes do not contain C-domains to form the peptide bonds, 2) the A domains do not charge amino acids (76). While another enzyme could, theoretically, form the peptide bond, the fact that NRPS-like enzymes do not use amino acids in the synthesis of their non-peptide products means that they can safely be ruled out of a role in malformins production.

Two of the NRPSs in *A. niger* have already been characterized, sidC and sidD coding for siderophore producing enzymes (77).

To identify which of the remaining NRPSs were the best candidate for malformins biosynthesis domain analysis and gene cluster analysis was employed. Since malformins are composed of five amino acid, two of which are cysteine, and each A domain is specific to one type of amino acid, the malformins NRPS must contain at least four A domains. Of the remaining NRPSs, five have four A domains and three have five A domains. These can further be narrowed down by identifying if the NRPS is linear, iterative, or non-linear. NRPSs with the simple repeating domain order of A-T-C tend to be linear (35). NRPSs with repeating C domains tend to be iterative (35). NRPS with five A domains can be linear and allow for malformins synthesis, and those with four A domains must be iterative or non-linear. NRPSs with methyltransferases can also be omitted as there are no extra methyl groups on malformins. NRRL3_10912 and NRRL3_08891 are also omitted as they contain two C-C domain groups, meaning they are capable of iterating more than is required. This leaves 1 four domain NRPSs and 2 five domain NRPSs.

Since gene clusters are thought to be assembled so that multiple genes in a pathway can be co-regulated gene cluster analysis was done to look for other genes required for malformins synthesis. Since malformins contains a di-sulfide bond an enzyme capable of forming this bond is expected to be present. An analysis of different clusters showed a thioredoxin reductase adjacent to *NRRL3_08969*.

Recently, the knockout of an NRPS in *Aspergillus brasiliensis* (*Aspbr1_34020*) showed abolishment of malformins production, followed by its rescue when the NRPS was reintroduced (10). This NRPS has the greatest sequence homology with *NRRL3_08969* in *A. niger* with 82% identity.

The second cluster implicated, the NRRL3_07881/NRRL3_07884 cluster, does not contain an NRPS. Instead it contains a polyketide synthase and a NRPS-like gene, according to internal annotations. Neither of these two genes are expected to produce enzymes capable of synthesizing malformins, because as reasoned previously, malformins is a non-ribosomal peptide. Polyketide synthases produce polyketides (20) and NRPS-like enzymes do not produce peptides (76).

3.2.2 Malformins pathway analysis strategy

Previously, upregulation of a transcription factor, *NRRL3_07873*, resulted in a median increase of Malformin A1 production of approximately 83-fold, according to mass spectrometry peak area. These results were gained from 11 independent transformations where the *NRRL3_07873* transcription factor was randomly integrated into the *A. niger* genome with a GlaA promoter in unreported data from S.Palys. Each independent transformant showed different levels of Malformin A1 production, ranging from a peak area of $4.7X10^3$ to $2.2X10^7$, indicating that some uncontrolled variable is affecting malformin A1 levels. This uncontrolled variable could be related to the location in the genome into which the transcription factor was inserted.

In order to better control an experiment directed at understanding malformins regulation a malformins producing strain needed to be generated in which the location where transcription factor was inserted was known. The first step was to replicate the results of S.Palys with a random integration of the *NRRL3_07873* gene with GlaA promoter. Given that metabolomics data was available from the larger, ongoing, transcription factor overexpression project the malformins levels of all strains were investigated. This included the *NRRL3_07873* transcription factor that showed interesting results in the random integration strains. *NRRL3_08965.1* was of particular

interest given that it resides in the putative malformins cluster. All these strains are generated from the Δ*KusA*Δ*PyrG* NRRL 2270 *A. niger* parental strain.

The enzymes in the malformins pathway could then be characterized by knocking out genes of interest and looking for intermediates or abolishment of malformins. The results of Theobald *et al*. (10) could be repeated by knocking out the backbone enzyme, as was the original intention before their study was published. This could still be useful as the Theobald *et al*. study was done on *Aspergillus brasiliensis* and so this knockout would confirm the function of the *A. niger* homolog. Knock-out of the thioredoxin reductase gene *NRRL3_08968* could result in accumulation of a malformins intermediate without a disulfide bond.

3.2.3 Malformins Identification

Malformins is commonly seen in literature at three retention times, as seen in the Theobald *et al*. paper (10). These three retention times are a result of overlapping retention times of the nine forms of malformins, making them difficult to distinguish. What can be distinguished are the two different masses represented in malformins molecules. The smaller of the retention time peak at 12.95 minutes accounts for the 515 mass malformins while the other two account for the 529 mass malformins. This was determined by looking at the mass peaks present at each individual retention time. The presence of both masses adds to the probability that these are true malformins peaks. These masses distinguish themselves from noise due to isotopic sub-peaks and adducts. This cannot be said of two of the 12 negative controls which only contain a single mass peak in some of the samples, though given the relatively small size of these peaks it is not surprising the smaller sub-peaks cannot be seen.

Further corroborations of the identity of these compounds could be achieved in a number of ways. A standard of synthetically produced malformin could be purchased and run in the same chromatography column as the fungal metabolites. If the molecules are the same, then they should have the same retention time in any column that they are both run through. As of July 13, 2020, malformin C could be purchased from Enzo life sciences at 257.00 USD for 10 mg. While expensive for mass production the price is reasonable enough for the purposes of identification. It would become more difficult to do this for every malformin molecule as this would entail extra costs and some forms of malformin have yet to be synthesized. Identifying one peak as malformin would increase the likelihood of the other peaks also being malformin as it is not unusual for them to be produced together (25). As such running a standard may still be a useful step.

Nuclear magnetic resonance could be used to look at specific features on the molecule to corroborate its identity (78). The difficulty with this process is in the need to grow enough metabolite to produce a readable signal as this technique is not as sensitive as mass spectrometry. Tandem mass spectrometry can also be used to further fragment molecules into predictable masses. The mass spectrum of the fragmented compound can be compared to an existing spectrum from that compound in a database such as the MassBank of North America. This database was used to compare tandem mass spectrometry from the metabolome of the *NRRL3_07873* targeted integration strain, and the spectrum was a match with a malformin A2 spectra seen in the database. However, at the ionization energy used there were only a couple identifying peaks, reducing the overall certainty of the identification.

3.2.4 Malformins levels

Upregulation of the malformins gene cluster transcription factor *NRRL3_08965.1* did not produce the expected increase in malformins production that was seen, for instance, with

upregulation of the alkylcitric acid gene cluster transcription factor and its associated metabolites (67). The seemingly unrelated transcription factor, *NRRL3_07873*, produced better results but there are still questions that need to be addressed regarding why different strains produce different levels of malformins. If the goal is to maximize malformins production and understand its regulation the factors limiting its production should be discussed. In doing so the factor responsible for the different levels of malformins in the *NRRL3_07873* strains may be uncovered.

In the case of alkylcitric acid, overexpression of the gene cluster transcription factor *akcR* resulted in an increase of alkylcitric acid from a mg/L level to a g/L level, a roughly 200 fold increase in *A. niger* (67). While the mass yield of malformins in the *NRRL3_07873* overexpression strains has not been measured, malformins peak area was increased between 200 and 5000 fold, depending on the strain. At its highest, malformin peak intensity has only been observed to reach $2.5X10^6$ compared to the $1.9X10^7$ seen for hexylaconitic acid A, the most abundant of the alkyl citric acids. These numbers are not directly comparable as different molecules could have a different capacity to be ionized and detected by the mass spectrometer, but it does represent nearly an order of magnitude of difference (79). The Theobald paper, which overexpressed the putative malformins backbone enzyme in *A. brasiliensis*, reported a peak height of $2X10^6$. Since NRPSs are capable of cyclization the only other enzyme required would be the disulfide bond-forming enzyme. Given that no disulfide bond-forming enzyme was intentionally overexpressed in the Theobald paper it is possible that the di-sulfide bond forming enzyme remains the rate-limiting step, provided the supply of amino acids is not rate limiting. Our data shows basal expression of the putative disulfide bond-forming enzyme gene *NRRL3_08968*, annotated as a thioredoxin reductase, in the NRRL 2270 strain. If *A. brasiliensis* has similar levels of expression of its disulfide bond forming enzyme it is surprising only upregulating the backbone enzyme shows such high levels of malformins. Disulfide bonds do form spontaneously in vivo, but at rates orders of magnitude lower than with an enzyme (80). It is possible that upregulation of the backbone enzyme also resulted in upregulation of the disulfide bond forming enzyme through chromatin opening as there is a good candidate for the role adjacent to the backbone gene. This is true in *A. niger* as it is in *A. brasiliensis*, as there is a high degree of homology for this cluster in both strains. In addition, NRPSs would take longer to perform their multistep reactions than the reactions required for a single disulfide bind, likely meaning less of the disulfide-bond forming enzyme would be needed before it stops being rate-limiting. It has been reported recently that the upregulation of the *Aspergillus brasiliensis* malformins cluster thioredoxin reductase (mlfE) in *Aspergillus nidulans* did not affect disulfide bond formation in malformins. This is the homolog NRRL3_08968 so the correct gene may lie elsewhere (81). Another limit to production of malformins may be toxicity. As malformins has been reported to be toxic to *Aspergillus nidulans* there may be a point where organism undergoes enough stress to undue any gain made by pathway engineering (81). Rendsvig *et al*. reports that the transporters in the *Aspergillus brasiliensis* malformins cluster does reduce toxicity (81). What further increases in malformins production can be gained by modifying the pathway is difficult to say. If *NRRL3_07873* only upregulates *NRRL3_08969* it is possible that any effort to increase its expression alone will not further increase malformins production past this upper limit.

The increased level of malformins produced the *NRRL3_07873* random integration strain compared to the *GlaA* locus integration strain suggests an uncontrolled variable is affecting the expression of downstream genes. That is to say that for every random integration event there is an uncontrolled variable that is being affected differently, and the GlaA integration strain represents a single value of this variable that happens to result in lower malformins levels. Since

transcriptomics have shown that the GlaA insertion actively transcribes *NRRL3_07873* it is unlikely that the lower levels of malformins in this strain are due to compromised *NRRL3_07873* expression. There is the possibility of multiple insertion events in the random integration strain, which could result in increased transcription factor levels, but it is questionable if increased expression beyond what is seen in the targeted strain would improve malformins production. To investigate the hypothesis that transcription factor gene copy number could have an effect on malformins expression there are a few experiments that could be conducted. Copy number of the transcription factor gene in the random integration strain could be quantified by quantitative PCR. Expression levels of *NRRL3_07873* could be measured in the random integration strain to measure the effects of *NRRL3_07873* expression on malformins production more directly. Showing an increase in copy number or expression would reveal a possible reason for increased malformins expression, but not knowing the insertion loci leaves other confounding variables that could influence malformins production. In order resolve this a new strain could be constructed with multiple transcription factor integrations at known loci. The *A. niger* strain A1513, for instance, contains 4 glucoamylase loci. Using existing DNA constructs multiple strains with between 1 and 4 transcription factor integration events could be constructed.

Other reasons for this discrepancy have been proposed. It is possible integration of the glucoamylase promoter could affect nearby chromatin and alter gene expression levels of nearby genes. Using this principle, the glucoamylase promoter could be introduced in front of the transcription factors while they are in their proper clusters. Introducing the promoter to NRRL3_08965.1 would not be expected to increase malformins production, unless chromatin opening in the cluster allowed expression of cluster genes. The *NRRL3_08965.1* gene could be replaced with *NRRL3_07873* to test if this transcription factor enhanced this effect, demonstrating the superior regulatory control of *NRRL3_07873* over the cluster and the dependence of integration loci on malformins expression.

The random integration could have also directly disrupted nearby genes that act as negative regulators, but the chance of this occurring is low considering the size of the genome. *Aspergillus niger* contains 597 transcription factors (513 of which are annotated as fungal-specific) and 33,795,461 base pairs according to our internal data which may have been updated since time of last publication (54). The chances of a relevant gene being disrupted by one or a few integration events is unlikely. That multiple random integration strains were produced with high levels of malformins production makes the chance of a specific gene being the affected targeted even more unlikely. This is not to say negative regulators are not involved in malformins production, only that their disruption likely does not explain the difference between the random and specific integration strains. There is the possibility that if transcription factor expression is too high a negative regulator may cause negative feedback, which would implicate a negative regulator, explain the high levels of expression of *NRRL3_07873* with low levels of malformins production, and allow for different levels of malformin production with different integration events. *NRRL3_08974*, for instance, is a negative regulator of transcription factors, and while it is not in the defined malformins cluster it is just outside of it.

Malformins production in these strains have not been controlled to growth rate, so it is possible that some of this increased level is due to increased biomass, or a re-allocation of biomass to metabolite production. The more than 10-fold difference in Malformin levels between strains, however, does not reflect the possible differences in culture growth of cultures that are ostensibly the same size. The 96 well plates do not allow for large differences in mycelial mass once the mycelial plugs have grown into the wells. Spore count was controlled for, so differences in

malformins production do represent some difference between the strains, even if the increase is not directly due to upregulation of the backbone gene.

These experiments do not establish the causal nature of the link between the transcription factor and the putative malformins NRPS. To test whether NRRL3_07873 directly interacts with the *NRRL3_08969* promoter would require an experiment like Chromatin Immuno-Precipitation (ChIP) where the transcription factor would be cross-linked to the promoter in order to corroborate its interaction (82). It is possible that the system regulating malformins production includes multiple transcription factors acting in series or parallel, negative feedback loops with negative regulators, or some form of epigenetic regulation. Producing a strain that produces levels of malformins higher than a simple *NRRL3_08969* overexpression strains could point to other genes involved in malformins production. Such strains could also implicate new genes in malformins regulation. It is possible that malformins expression is best achieved with multiple transcription factors being overexpressed, such as *NRRL3_07873* and *NRR3_08965.1*. It is also possible that the effectiveness of a given transcription factor is reduced in this strain, and an orthologous transcription factor in another strain would produce better results. This could happen by spontaneous mutations in laboratory strains deactivating energetically expensive positive regulators of secondary metabolite production, as has been observed in the past (1).

There is the possibility of repressors acting to repress malformins production. One way of discovering new genes is by creating a selective pressure towards mutations in the genes of interest. In the case of a repressor, production of a necessary nutrient could be linked to expression of a gene the repressor is supposed to repress. Provided that the promoter is capable of transcription in the absence of the repressor the *A. niger* genome could be mutagenized and the resulting surviving strains could be analyzed to look for common mutations. The only survivors on selective media should be those which have removed the functionality of the repressor (or mutated the promoter to avoid the repressor). In the context of *NRRL3_07873*, the selective gene would be placed in front of a putative downstream gene promoter such as *NRRL3_08969* in the low malformin producing *NRRL3_07873* overexpression strain in an attempt to alleviate the repression. Alternatively, this could be done in a wild type strain to see if the gene has any basal level of expression absent the repression.

There are master regulators; regulators whose expression is implicated in the expression of many secondary metabolites. LaeA is a more well-known example in Aspergilli (43). Recently McrA has emerged as master repressor of secondary metabolite production (41). LaeA is almost universally a positive regulator of secondary metabolite production, so a knockout should not likely facilitate malformins production, though the possibility remains that it could negatively regulate genes which require some stimulus to be expressed (43). In these cases, the negative regulation would not be observed in the wildtype strains where expression is already basal.

No malformins was observed in a strain where *LaeA* was deleted. Growing the parental strain in the dark did produce malformins more often than in the other growth conditions tested, but the peaks were at very low levels (the largest being $1.23X10^4$) making strong conclusions difficult to draw. Since McrA was demonstrated to be a negative regulator a knock-out could reveal new metabolites or increase metabolite levels in cases like that of malformins where the metabolite is already observed. Elimination of repression may not be enough however, as some genes may require induction as well. This may be the case with the low level of transcription in otherwise silent genes in the *McrA* knockout strain produced by Oakley *et al* (41).

3.2.5 External stimuli and environmental conditions for malformins production

In order to investigate alternate regulatory pathways for malformins these strains were subjected to complex mixtures of biomass and stress conditions. None of the conditions tested resulted in increased malformins production. No malformins production was observed in these conditions, except in the cultures grown in the dark. This dark response could be involvement of the velvet complex, but these levels are too low to draw any conclusions.

Minimal media+1% agar media with and without 5% salt supplementation were also tested on the parental strain and several *NRRL3_07873* overexpression strains. This was an attempt to screen for inducing conditions across strains in case the inducing conditions required transcription factor upregulation as well. These resulted in lower levels of malformins than the targeted integration strain grown in 96-well plate with minimal media+1% maltose liquid media. As such these conditions are not expected to trigger synergistic pathways for malformins production.

One possible way to induce malformins production is to attempt to expose it to organisms it may have evolved with in nature, or to organisms that produce similar compounds to those other organisms. For instance, Jomori *et al*. induced malformins production using a coculture with *Mycobacterium smegmatis* in *A. niger* (83).

3.2.6 Expression analysis by RNAseq

RNAseq from RNA taken from a two-hour transfer culture did not show upregulation of cluster genes from either the *NRRL3_07881/NRRL3_07884* or *NRRL3_08969* cluster. This lack of expression could be because a different cluster is involved, but no other NRPSs were observed to be upregulated. If any of these genes are required for malformins production, they are expected to increase in expression in strains that produce more malformins. The caveat to this is that the genes have to be upregulated at some point between induction of the transcription factor genes and production of the metabolites, not at any given point. In 96 well plate cultures peak malformins levels are at 5 days of growth on inducing media. This is well after two hours, but the transfer culture conditions are different and so malformins production may not take as long to start. To this end a time course of the transfer culture was conducted to look for active malformins production.

3.2.7 Transfer culture for RNAseq

The transfer culture was first conducted with the *NRRL3_07873* random integration strain because shaking cultures tend to produce less metabolite and the random integration strain was expected to give the clearest signal. Time points were selected at 2, 8, and 24 hours. The 2 hour and 8 hour time points produced no observable levels of malformins. Malformins did appear at 24 hours but lacking further data points it could not be concluded whether this was trending up or down. The experiment was repeated with the targeted integration strain that had been submitted to RNAseq, only this time the 8 hour time point was replaced with a 48 hour time point. Results showed malformins levels decreased after the 24 hour time point, and so production is expected to be lower after this time point as well. While there is typically a lag time between upregulation of a gene and production of the metabolite 6 to 22 hours is outside the norm. Takahashi *et al* looks at a variety of cellular processes and shows gene expression to metabolite lag times of 10 to 90 minutes (84). Nakabayashi *et al*. observed flavanol gene expression upregulation at 3 hours with metabolite being observed 3 hours later (85). It is therefore not unreasonable to conclude that the cluster might be upregulated at a later time-point.

3.3 Master negative regulator *McrA* depletion strain

It was hypothesized that a negative regulator could be responsible for counteracting the effects of gene cluster transcription factor expression. If the negative regulator either binds the transcription factor directly, or to a downstream target of the transcription factor, then less secondary metabolite might end up being produced. Disrupting or knocking out the regulator would, therefore, be expected to increase metabolite production. *McrA*, as a master negative regulator of secondary metabolite production, is more likely to affect any given secondary metabolite gene cluster than a randomly selected negative regulator.

The depletion mutant of *McrA* in the NRRL 2270 background did result in a somewhat different metabolite profile; at least one that is more different than the difference between runs. 30.87 % of the variance between all samples come from entities which were differentially expressed between strains on two different dates. This compares to 18.65% of the variance coming from entities differentially expressed between runs. What proportion of entities are expected to be differentially expressed would depend on how many pathways McrA regulates. It is sufficient for the purposes of demonstrating that these strains are different to show that the variation between strains is overrepresented, though the exact confidence of this conclusion given the proportion of variance of each component is not calculated in this study. It can also be noted that principle component analysis may be unable to separate strains even if they are different if the gene in question is responsible for fewer entities than typically vary between samples. This technique was used to analyze McrA specifically because McrA is a master regulator.

Making analysis of which metabolites are regulated by this regulator difficult is the noisiness of the data. The more variance there is between runs, and between samples, the less certain it is that any given metabolite is the result of *McrA* depletion as opposed to a noise. The more metabolomics is run on the strains, and the more a given compound is seen upregulated or downregulated, the more certain it is that the compound is regulated by the gene in question. One potentially interesting compound present in all six samples of the *McrA* depletion strain but not the parental strain is a compound with the same mass as Asticolorin A, within 10 ppm. Asticolorin A is a xanthene that has only been observed in *Aspergillus multicolo* (86). No genes have been associated with its production. Looking for upregulated clusters in this strain could potentially provide candidates for Asticolorin A production. Other compounds identified by software according to mass are listed in table 12. Some of these annotations are more questionable than others, such as Miraxanthin-I, a yellow-plant pigment, which is less likely to be produced in a fungus than known fungal metabolites. This could mean that the peak instead represents a novel compound that happens to share the same or approximately the same mass (within 10 ppm).

Malformins was not seen in the *McrA* depletion strain. This could be because malformins production requires a positive regulator in addition to removal of a negative regulator, or McrA may not be involved. Upregulation of gene cluster transcription factors in the *McrA* depletion strain would be needed to investigate this question.

It is possible that the removed portion of the *McrA* gene may have actually enhanced its effect. With the promoter region removed and start codons in frame with the DNA binding domain still intact, it is possible that negative regulation of the negative regulator has been relived. *McrA*'s transcript has an unusually long 5' non-coding region indicating regulatory complexity. On the other hand, removal of the promoter may have also removed the genes ability to recruit RNA polymerase (87). To resolve this question, RNA expression or proteomics would be needed to look for expression of the remaining gene fragment.

The purpose of *McrA* depletion in the NRRL 2270 strain was to test its effects on malformins production in the same background strain as the transcription factor overexpression strains. It was hypothesized that *McrA*'s role in regulating secondary metabolism could be limiting the expression of malformins in the putative malformins' TF overexpressing strains. In *A. nidulans* ChIP has shown that McrA is involved in regulation of secondary metabolites in many ways. McrA binding sites, for instance, were found upstream of gene cluster transcription factors, negatively regulating them (87). This effect would be redundant with our glucoamylase promoter driven transcription factor expression mutants as these mutants not only have high levels of transcription factor expression, but the *GlaA* promoter may not be targeted by McrA. These binding sites were also found on biosynthetic genes within the cluster, meaning even with upregulated transcription factors the enzymes themselves may experience repression. McrA also interacts with other binding sites upstream of other regulator genes, such as *LaeA*, *McrA* itself, and non-coding RNAs surrounding it (87).

In terms of epigenetics, while replacing the glucoamylase promoter of the cluster transcription factor may alleviate some of the local heterochromatin, the presence of McrA at other cluster genes may nullify this effect. On top of this, McrA does affect other genes which could be involved indirectly . The methyltransferase *llmG* is a positive regulator of secondary metabolism, speculated to be due to its ability to open up chromatin by chromatin methylation. McrA has been shown to be a negative regulator of *llmG*, allowing an alternative route to cluster downregulation beyond directly binding the clusters (88).

3.4 Conclusion and suggestions for future work

Overexpression of the *NRRL3_00042* biosynthetic gene cluster transcription factor gene resulted in the appearance of three peaks of unidentified mass: 425.1368, 284.1277, and 409.1383. The overexpression also resulted in the mycelia having a strong yellow color and a reduced growth rate. Deletion of the backbone enzyme gene *NRRL3_00036* resulted in the partial restoration of the parental strain phenotype and the disappearance of the listed mass peaks.

The identity of these peaks still needs to be investigated by a technique to resolve the compounds structure, such as NMR. Potential bioactive effects should be screened to determine clinical relevance. The involvement of tailoring enzymes in the cluster could also be investigated and a platform for the production of the compounds could be developed.

Overexpression of the *NRRL3_08965.1* biosynthetic gene cluster transcription factor did not result in upregulation of its cluster. The overexpression of the NRRL3_07873 biosynthetic gene cluster transcription factor did show the greatest increase in malformins production, the compounds that have been bioinformatically associated to the NRRL3_08969 cluster.

The upregulation of the *NRRL3_08969* cluster in strains with increased malformins production still needs to be demonstrated by taking RNA at a more appropriate timepoint. The involvement of cluster enzymes, namely the NRRL3_08969 NRPS backbone enzyme and the NRRL3_08968 thioredoxin reductase, still needs to investigated. The regulation of the cluster could be further investigated by introducing a promoter containing AngCP binding sites to the cluster in an effort to open up cluster chromatin. Comparing this effect in strains with and without *NRRL3_07873* in the glucoamylase locus could implicate epigenetic regulation as the uncontrolled variable altering malformins production levels. Further, transcription factors such as NRRL3_07873 could be upregulated in the McrA depletion strain to test for increased metabolite production in the absence repression by McrA.

Materials and Methods

4.1 Strains

Aspergillus niger NRRL 2270 Δ*KusA*Δ*PyrG* is the parental strain used in this project for the generation of mutant strains by gene replacement and gene deletion. *A. niger* NRRL 3 was the reference strain for bioinformatics and genomics analysis. Gene names refer to the NRRL 3 annotation (i.e. *NRRL3_00036* encodes an NRPS). *NRRL3_00042* encodes the transcription factor located inside the *NRRL3_00036* cluster, *NRRL3_08965.1* is the transcription factor located inside the *NRRL3_08969* cluster, and *NRRL3_07873* is the transcription factor located inside the *NRRL3_07881*/*NRRL3_07888* cluster.

4.2 Polymerase Chain Reaction (PCR)

PCR reactions are done at varying volumes. The ratio of ingredients are as follows: 4 µl 5X HF buffer, 0.4 µl 10mM dNTP, 1 µl forward and reverse 100 mM primer, 0.1 µl phusion protein, and a total of 13.5 µl for the remaining ddH_2O and DNA template, with volumes shared between them depending on DNA template concentration.

The exact PCR cycling varies by length of desired amplicon and type of template DNA. PCRs begin with an initial denaturation step, which is 1 minute for plasmids and 5 minutes for genomic DNA. Next, 30 cycles of denaturation, annealing, and amplification are performed. Denaturation lasts 15-20 seconds, annealing lasts 15-30 seconds, and amplification lasts 30 seconds per 1 kb of DNA to be amplified. A final extension step lasts between 7 and 10 minutes.

4.3 Transformation Constructs

Two different transformation methods were used in order to generate mutant strains. One inserted genes randomly into the *A. niger* genome and the other used a CRISPR/Cas9 system to make double stranded cuts at specific sites within the genome and a homologous repair template to repair the cut as desired. Both utilized plasmids to contain and deliver the desired genetic information. Sequence information for plasmids was stored using the Clone Manager software, which also allowed for simulation of restriction enzyme cuts.

Targeted integration and deletions used ANEp8 plasmids defined according to Song *et al.* (89). When the guide RNA (gRNA) insert is included, ANEp8 are 15928 base pair plasmids containing the following features: A beta-lactamase (*bla*) gene encoding for resistance to beta-lactam family antibiotics (such as ampicillin), AMA1 sequence for plasmid replication, pyrG gene for auxotrophic selection, a Cas9 gene with PKiA promoter and GlaA terminator, and a gRNA.

The bla gene is used for selection of positively transformed *E.coli* colonies by growing in media with ampicillin. The AMA1 sequence allows for higher efficiency of transformation into aspergilli due to presence of mobile Aspergillus transformation enhancers and ARS-consensus sequence. The pyrG gene encodes orotidine 5'-phosphate decarboxylase, which is responsible for catalyzing the production of uridine monophosphate. Strains lacking the *pyrG* gene must be supplemented with uridine or be rescued. Cas9 is an enzyme used in bacteria immune response, which can catalyze double stranded cuts at specific sequences complementary to provided gRNA. Cuts occur at the after the 17 base pair targeted by the gRNA, three base pairs before the required PAM site, defined as a sequence of three nucleotides with the code NGG.

ANEp8 plasmids are generated by inserting a 359 base pair insert containing the desired gRNA. Inserts are created by amplifying two separate DNA fragments by PCR. Each amplification has one primer with a 20 bp overhang corresponding to the desired gRNA. This overhang allows

overlap PCR, where the overhangs of each fragment bind complementarily, and outside primers amplify the whole length of the two amplicons, creating a single insert.

The insert is incorporated into existing ANEp8 plasmid by ligation independent cloning (LIC). The "empty", meaning without insert" plasmid is cut with SwaI restriction enzyme. This linearizes the plasmid and allows digestion of the 3' end with T4 DNA polymerase. Circularization is prevented by treating the digested plasmid with calf-intestinal alkaline phosphatase (CIP). T4 polymerase is able to remove 3' nucleotides when those nucleotides are not present in solution. This means the enzyme will digest the DNA up until it reaches a nucleotide with which the solution is supplemented. Due to the design of this stretch of DNA, supplementing insert digestion with dCTP and the linear plasmid with dGTP creates complementary overhangs. The insert and plasmid are annealed together by heating the insert and plasmid together to 60°C, then gradually cooling the mixture to 4°C. The annealed plasmid is then transformed into *E.coli*. *E.coli* ligates the plasmid with the insert and propagates the construct.

Random integration utilized ANIp7 plasmids. These plasmids contain the same *bla* and *pyrG* genes found in ANEp8. Instead of a *Cas9* gene ANIp7 contains the glucoamylase promoter and terminator, with a PsiI restriction site in between the two features to allow integration of a transcription factor gene by LIC. Plasmid size before inclusion of the gene is 5642 bps. The plasmid lacks an origin of replication, meaning it must insert itself into the genome to be passed on with stability.

4.4 Mini Prep – Extraction of plasmids from *E.coli*

Transformed *E.coli* colonies from the ANEp8 transformation, grown on Lysogeny Broth+amp agar plates, are picked with toothpicks and placed into 7.5 mL liquid Lysogeny Broth+amp. These cultures are grown overnight at 37°C, at 220 rpm. Cultures are then spun down and undergo mini-prep with a Bio-Basic molecular biology kit according to the "Protocol for Purification of Low Copy Plasmid DNA" in the "EZ-10 Spin Column Plasmid DNA Miniprep Kit" manual (BS614/414/413, Ver.2017AM1, Rev.20.06.2017). Collected plasmids are quality controlled with a pair of primers spanning either side of the insert insertion site.

4.5 Guide RNA and repair template design

In *A. niger* the *KusA* gene is responsible for non-homologous end joining. Working with a Δ*KusA A. niger* strain means that any cuts to the genome must be repaired through homologous recombination. Due to this, any cuts not repaired with a homologous template will lead to cell death. In this way, it is expected that only colonies with the desired modification will grow.

Guide RNAs are selected according to cutting efficiency within the region that will be homologously recombined. In addition, gRNAs with off target binding sites are omitted. Geneious software was used to look for guide RNA targets, score their efficiency, and look for off target sites. For knockouts, repair templates are designed with 30 or 45 base pairs on either side of the region to be removed. Each of these arms are selected to have an equal annealing temperature to their counterpart. Higher binding affinity is expected to result in more efficient transformation.

Gene insertions are done by amplifying the desired sequence from an existing parental strain genome and adding homologous regions to either end that are complementary to where the sequence will be inserted. In the case where non-native promoters must be added to a gene overlap PCR is used, or, if the repair template is being propagated on a plasmid, the gene can be inserted into a plasmid that already contains the desired promoter.

For upregulating transcription factors in a high throughput method, a plasmid had been designed that already incorporates the glucoamylase promoter and terminator. This allows the amplification and insertion of any number of genes. Using this method limits homologous recombination to the glucoamylase locus, however, as adding homologous regions for new loci would require a new plasmid.

Primers with 60 base pair overhang homologous regions can be used, but these have lower transformation efficiency compared to plasmids, and require large PCR volumes to reach significant yield.

4.6 Transformation

Incorporation of these DNA constructs into *A. niger* cells requires permeabilization of the cell membrane. Typically, *A. niger* contains a polysaccharide cell wall, which gives the mycelia its shape and holds the cells together. This also gives the plasma membrane extra structural integrity. In order to facilitate permeabilization the membrane enzymes were used to digest the cell wall, leaving single, round, cells called protoplasts.

Protoplasts are delicate, so rough handling and over digestion can reduce protoplast yield. Since transformation efficiency is limited, a high number of protoplasts are needed to obtain transformed colonies. This means enzyme concentration and digestion time must be balanced. Further, the optimal protoplasting conditions vary by strain.

The NRRL 2270 strain in particular is resistant to protoplasting, relative to other *A. niger* strains such as PY11 and A1513. In order to maximize protoplast formation without running the digestion for too long several techniques were employed. A primary 25 mL seed culture was grown for 24 hours before 5 mL was transferred to a 100 mL culture. This culture was grown overnight for 15 hours. This transfer culture protocol is used to produce fresh, young mycelia. This allows for a high mass of mycelia that has not had time to grow a tougher cell wall. A marble was included in both these cultures to prevent clumping of mycelia and allow more surface area for enzymes to digest the cell walls. Both cultures are grown at 32°C, at 220 rpm, in CM media (0.056 M D-Glucose (Bioshop), 0.2% peptone (Bioshop), 0.1% Yeast Extract (Bioshop), 0.1% Casamino Acids (Bioshop), 0.01 M L-Uracil (Bioshop), 0.07 M $NaNO_3$, 0.007M KCl, 0.6 M KH_2PO_4, 0.6 M K_2HPO_4).

Mycelia was collected by filtering over miracloth and washing with 0.6M $MgSO_4$. One gram of novozyme enzymes is added to 100 mL OM solution (32 ml OM1 (0.005 M NaH_2PO_4, 0.0262 M Na_2HPO_4), 50 ml OM2 (2.4M $MgSO_4 \cdot 7H_2O$), 18 ml ddH_2O). The mycelia was added to the OM solution, vortexed to reduce clumping, and separated into four 250 mL glass flasks. Flasks were placed at 37°C at 100 rpm for 100 minutes. Protoplasts were then transferred by serological pipette underneath an ice-cold 20 ml layer of TB (0.6M sorbitol, 0.1M Tris-HCl at pH 7.5) in a 50 ml tube. Tubes were centrifuged for 40 minutes, 4°C, at 4000 rpm. This separated the solution into two phases with protoplasts floating between them. The middle phase was taken by serological pipette and pipetted underneath a 40 ml volume of SC solution (1 M sorbitol, 0.05 M $CaCl_2 \cdot 2H_2O$) in 50 ml tubes. These tubes were centrifuged at 3000 rpm, 4°C, for 10 minutes to pellet protoplasts. Supernatant is removed and pellets were resuspended in order to achieve desired protoplast concentration.

In order to introduce DNA into the protoplasts the cell membranes must be permeabilized. A 40 µl volume of protoplasts were taken per transformation to be performed. This includes: all transformations, negative controls with no DNA added and therefor no pyrG for survival, negative controls with just the ANEp8 plasmid which are expected to be lethal due to DNA cuts in the

Δ*KusA* strain, positive control with an ANEp8 plasmid without gRNA, and positive control where the protoplasts are plated on media supplemented with uracil. To these protoplasts a maximum of 10 µl of DNA was added, typically 1 µg of ANEp8 plasmid and 20 µg of repair template. Next, 4 µl of ATA and 20 µl of 20 % PEG (freshly prepared from 60% PEG in SC) were added. Mixing was done by gently pipetting and inverting the transformation tube. The tubes were then left to sit at room temperature for 10 minutes before 300 µl of 60% PEG is added. After being mixed by pipetting and inverting the tubes were left to sit at room temperature for 20 minutes. Next, 1 mL of sorbitol is added, and the transformation is mixed by pipetting. These tubes were centrifuged at 16,000 *g* for 4 minutes in order to pellet protoplasts. The protoplasts were then suspended in 200 µl sorbitol for plating on SRM plates.

Transformation plates were left at 30°C until spores begin to form (around 4 days). Spores were then toothpick transferred into 1 mL CM in 5 mL tubes in order to grow mycelia overnight at 32°C for genomic DNA extraction. Mycelia is spun down at 16,000 rpm for 5 minutes and supernatant is removed. An 800 µl volume of DNA extraction buffer (0.173 M SDS, 0.001 M Tris, 0.127 M EDTA) was added to each tube with 250 µl glass beads. Tubes were vortexed, then bead-beat with an MP Biomedical FastPrep-24 Classic Instrument. 10 µg RNAse A was added to the lysed mycelia and incubated for 30 minutes at 37°C. Tubes were spun down at 16,000 *g* to reduce bubbles from the DNA extraction buffer. DNA extraction then occurs by adding 800 µl phenol:chloroform:isoamylalcohol and vortexing for 1 minute. Tubes were spun down at 10,000 rpm for 10 minutes. A 700 µl volume of the upper phase is then transferred to a new tube where 700 µl of phenol:chloroform:isoamylalcohol is added. Tubes were spun down at 10,000 rpm for 10 minutes. A 500 µl volume of the upper phase is then transferred to a new tube where 500 A 700 µl volume of the upper phase is then transferred to a new tube of isopropanol and 50 µl 3M sodium acetate, in order to precipitate DNA. Samples were vortexed and left at -20°C for 30 minutes. Tubes were then centrifuged at 16,000 *g*, 4°C, for 30 minutes in order to pellet precipitated DNA. Supernatant is removed and the pellet is washed twice with 250 µl volumes of cold 70% ethanol. Pellets were suspended in 45 µl warm ddH_2O.

In order to ensure proper transformation, the region of the genome being transformed is checked. For deletions spanning primers were used and the deleted region is measured by the difference in band size vs the control. The same technique is done with gene insertions, except in the case where the new sequence is the same size as the sequence being replaced. In this case, restriction enzymes were used that cut the amplicon of the inserted sequence and yield different DNA band fragments than the parental strain.

4.7 Growth conditions

For analyzing the metabolomes of strains overexpressing genes under the glucoamylase locus strains were grown in 190 µl of inducing medium containing maltose, namely minimal media +1% maltose (0.029 M D-Maltose, 0.0112 M $MgSO_4$, 0.07 M $NaNO_3$, 0.007 M KCl, 0.006 M KH_2PO_4, 0.006 M K_2HPO_4, 0.1% Hunter Trace Elements) and MMJ (0.438 M D-Maltose, 0.045 M $MgSO_4$, 0.4% Hunter Trace Elements, 0.28 M $NaNO_3$, 0.0028 M KCl, 0.024 M KH_2PO_4, 0.024 M K_2HPO_4). Typically, these cultures reach optimal growth at 5 days of growth when incubated at 30°C. With some strains time courses are used to find the optimal incubation time for production of the compound of interest.

For testing inducing molecules and conditions for malformins production, 24 well plates were used. Various media were used: MMJ or Minimal media+1% maltose supplemented with 1% wheat and/or 5% NaCl, and Yeast Malt Extract (0.3% yeast extract, 0.3% malt extract, 0.5%

peptone, 0.056 M D-glucose, pH adjusted to 6.00 with HCl). Different conditions were also tested: growing cultures in total darkness by wrapping plates in aluminum foil, growing fungi in 24-well polystyrene plates on a 30 °C hot plate by a window receiving 14 hours of sunlight per day, growing cultures at 30°C and 37°C, and growing cultures at 7 or 10 days.

4.8 Growing cultures for extracting RNA and RNA sequencing

For collecting RNAseq data liquid shaking cultures were used. Pre-cultures were grown in 100 mL CM +2% fructose in 1 L flasks for 16 hours at 30°C and 220 rpm. Mycelia was collected by filtration with miracloth and washed with sterile ddH_2O. A half tablespoon was transferred to 50 mL MM+1% cultures in 250 mL flasks. Cultures were grown for 2 hours at 30°C and 220 rpm. Mycelia was collected with miracloth and moisture was removed by folding the miracloth over and squeezing with paper towel. Samples were frozen with liquid nitrogen and the mycelia was grinded with mortar and pestle. Mycelia was added to trizol in 100 mg mycelia per 1 ml trizol portions. Samples were vortexed 30 seconds and incubated a5 5 minutes at room temperature. Samples were centrifuged at 10,000 rpm and 4°C for 15 minutes. Supernatant was transferred into a new tube and 200 µl chloroform was added. Samples were shaken by hand for 14 seconds and incubated at room temperature for 2.5 minutes. Samples were shaken by hand and then centrifuged for 10,000 rpm at 4°C for 15 minutes. The top, aqueous, phase was transferred to a new tube and 500 µl isopropanol was added to precipitate RNA. Samples were vortexed and left to incubate at room temperature for 10 minutes. Samples were centrifuged at 10,000 rpm and 4°C for 10 minutes. Supernatant was removed and pellets were washed with -20°C 1 ml 75% ethanol. Samples were centrifuged 5 minutes at 7,000 rpm and 4°C. The wash was repeated a second time. Pellet was dried in a roto-vac. The pellet was resuspended in 30 µl of nuclease free H_2O. The pellet was then dissolved by leaving on ice for 15 to 30 minutes.

Since there may be some lag time between the induction of a transcription factor and the expression of cluster genes it targets various secondary culture times were tested to look for peak metabolite production. A time course of transfer culture secondary cultures was done for the *NRRL3_00042* overexpression strain and the *NRRL3_07873* overexpression strain.

4.9 Mass Spectrometry

Metabolomes were analyzed by mass spectrometry. To prepare samples for mass spectrometry media from fungal culture wells was taken and spun down at 16,000 *g* for 30 minutes to pellet biological material. Supernatant was taken and protein was precipitated with 2:1 cold methanol. Samples were placed at -20°C for 30 minutes. Samples were then centrifuged at 16,000 *g* for 30 minutes to pellet precipitated protein. The top 50 µl of supernatant was taken and distributed into wells of a 96 well plate with 50 µl of 0.1% formic acid in ddH_2O.

Analysis of NRRL3_00036 cluster strains and malformins producing strains was done using a 7-Tesla Finnigan LTQ-FT mass spectrometer (Thermo Electron Corporation, San Jose, CA). Injection of 10 µl of each sample into a Kinetex 150 x 2.1 mm, 5 µm, C18 column (Phenomenex, Torrence, CA, USA) was done in order to separate compounds with an Agilent 1260 Infinity II HPLC system (Agilent technologies, Santa Clara, CA, USA). The solvents Solvent A (0.1 % formic acid in water) and Solvent B (0.1% formic acid in acetonitrile) were used to generate a reversed-phase separation gradient. A solvent flow rate of 250 µl/min was set with the following gradient conditions: 3% B isocratic for 1 min, increased to 80% B over 10 min, increased to 95% B in 0.1 min, maintained at 95% for 1 min, decreased to 3% B in 0.1 min and kept at 3% B for 4.8 min. Column eluate underwent electrospray ionization-mass spectrometry in the LTQ-

FT with an ionization voltage of 4900 V in positive mode and 3700 V in negative mode. Scan range spanned from 100 to 1400 m/z at 100000 resolution at 200 m/z.

Analysis of the NRRL3_03076 depletion strain was done using a Agilent 6560 ion mobility Q-TOF mass spectrometer equipped with a Jetstream electrospray source. Injection of 10 µl of each sample into a Kinetex 150 x 2.1 mm, 5 µm, C18 column (Phenomenex, Torrence, CA, USA) was done in order to separate compounds with an Agilent 1260 Infinity II HPLC system (Agilent technologies, Santa Clara, CA, USA). The solvents Solvent A (0.1 % formic acid in water) and Solvent B (0.1% formic acid in acetonitrile) were used to generate a reversed-phase separation gradient. A solvent flow rate of 450 µl/min was set with the following gradient conditions: 3% B to 80% B over 8 minutes. This was followed by an isocratic step with a solvent flow rate of 450 µl/min set with the following gradient conditions: 85% B for 1.5 minutes. The final re-equilibration step was performed at a flow rate of 450 µl/min with the following gradient conditions: 3% B for 3.5 min. Eluate was passed to the mass spectrometer. The negative and positive polarity acquisitions had a full scan range between 100 to 1700 m/z at 2 Hz out using 2 GHz extended dynamic mode range at 30000 resolution.

Data processing was done using Profinder for the McrA strain, followed by data analysis was done using Mass Profiler Professional. Xcalibur software was used to generate gene cluster strain chromatograms. Chromatogram labels were edited for legibility.

Before analysis by Mass Profiler Professional data is processed with Profinder and converted into .CEF files that can be opened in Mass Profiler Professional. In Profinder, the feature extraction algorithm for small molecules/peptides is selected. Batch recursive feature extraction was done with the following parameters: only peaks within a retention time window between 0.5 and 11, and peaks with heights of at least 1000, were extracted, the ion species allowed were: +H, +Na, +NH4, -H, +HCOO, and H_2O, assigned charge rates were limited to a range of 1 to 2. Compound filters were set with a compound ion count threshold set to two or more ions. Binning and alignment tolerances were set to a retention time window of 2.00% + 0.10 minutes with a mass window of 20 ppm + 2.00 millidaltons. Post processing filters used were: setting the MFE score minimum a minimum of 70 in at least 2 files of at least one sample group. EIC tolerance was set with expansion values of ± 35.0 ppm and extracted ion chromatogram range limited to an expected retention time of ± 0.10 minutes. EIC peak integration was done with Agile 2. Smoothing was done with the gaussian function with a function width of 9 points and a gaussian width of 5. Peak filters were filtered with a minimum absolute height of 5000 counts. Chromatogram data format was set to centroid. Peak spectra were included where average scans were greater than 10% peak height, excluding time of flight spectra when above 10% saturation. Centroiding is done with a maximum spike width of 2 and a required valley of 0.70. Tgt score minimum was set to 70 in at least 2 files of at least one sample group.

Once .CEF files are generated files were imported into Mass Profiler Professional and assigned parameter values. Since data was already filtered Mass Profiler Professional filters can be set with little stringency to allow al data to be used. Principle component analysis was done using the Mass Profiler Professional software.

References

1. Yan Q, Lopes LD, Shaffer BT, Kidarsa TA, Vining O, Philmus B, Song C, Stockwell VO, Raaijmakers JM, McPhail KL, Andreote FD, Chang JH, Loper JE. 2018. Secondary Metabolism and Interspecific Competition Affect Accumulation of Spontaneous Mutants in the GacS-GacA Regulatory System in Pseudomonas protegens. mBio mBio 9:e01845-17.
2. Schrey S, Erkenbrack E, Früh E, Fengler S, Hommel K, Horlacher N, Schulz D, Ecke M, Kulik A, Fiedler H-P, Hampp R, Tarkka M. 2012. Production of fungal and bacterial growth modulating secondary metabolites is widespread among mycorrhiza-associated streptomycetes. BMC Microbiol 12:164.
3. Künzler M. 2018. How fungi defend themselves against microbial competitors and animal predators. PLOS Pathog 14:e1007184.
4. Eidem HR, McGary KL, Rokas A. 2015. Shared Selective Pressures on Fungal and Human Metabolic Pathways Lead to Divergent yet Analogous Genetic Responses. Mol Biol Evol 32(6):1449–1455.
5. Daley DK, Brown KJ, Badal S. 2017. Chapter 20 - Fungal Metabolites, p. 413–421. *In* Badal, S, Delgoda, R (eds.), Pharmacognosy. Academic Press, Boston.
6. Dembitsky VM. 2014. Naturally occurring bioactive Cyclobutane-containing (CBC) alkaloids in fungi, fungal endophytes, and plants. Phytomedicine 21(12):1559–1581.
7. Hulvová H, Galuszka P, Frébortová J, Frébort I. 2013. Parasitic fungus Claviceps as a source for biotechnological production of ergot alkaloids. Prague Symp 2011 31(1):79–89.
8. de Groot ANJA, van Dongen PWJ, Vree TB, Hekster YA, van Roosmalen J. 1998. Ergot Alkaloids: Current Status and Review of Clinical Pharmacology and Therapeutic Use Compared with Other Oxytocics in Obstetrics and Gynaecology. Drugs 56(4):523–535.
9. Anderegg RJ, Biemann K, Buechi G, Cushman M. 1976. Malformin C, a new metabolite of Aspergillus niger. J Am Chem Soc 98(11):3365–3370.
10. Theobald S, Vesth TC, Rendsvig JK, Nielsen KF, Riley R, de Abreu LM, Salamov A, Frisvad JC, Larsen TO, Andersen MR, Hoof JB. 2018. Uncovering secondary metabolite evolution and biosynthesis using gene cluster networks and genetic dereplication. Sci Rep 8:17957.
11. Wang J, Jiang Z, Lam W, Gullen EA, Yu Z, Wei Y, Wang L, Zeiss C, Beck A, Cheng E-C, Wu C, Cheng Y-C, Zhang Y. 2015. Study of Malformin C, a Fungal Source Cyclic Pentapeptide, as an Anti-Cancer Drug. PloS One 10:e0140069–e0140069.
12. Tan Q-W, Gao F-L, Wang F-R, Chen Q-J. 2015. Anti-TMV activity of malformin A1, a cyclic penta-peptide produced by an endophytic fungus Aspergillus tubingensis FJBJ11. Int J Mol Sci 16(3):5750–5761.
13. Nguyen KT, Ritz D, Gu J-Q, Alexander D, Chu M, Miao V, Brian P, Baltz RH. 2006. Combinatorial biosynthesis of novel antibiotics related to daptomycin. Proc Natl Acad Sci U S A, 103(46):17462–17467.
14. Felnagle EA, Jackson EE, Chan YA, Podevels AM, Berti AD, McMahon MD, Thomas MG. 2008. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. Mol Pharm, 5(2):191–211.
15. Sasaki S, Hashimoto R, Kiuchi M, Inoue K, Ikumoto T, Hirose R, Chiba K, Hoshino Y, Okumoto T, Fujita T. 1994. Fungal metabolites. Part 14. Novel potent immunosuppressants, mycestericins, produced by Mycelia sterilia. J Antibiot (Tokyo) 47(4):420–433.

16. Vardanyan R, Hruby V. 2016. Chapter 33 - Antifungal Drugs, p. 677–686. *In* Vardanyan, R, Hruby, V (eds.), Synthesis of Best-Seller Drugs. Academic Press, Boston.
17. Banani H, Marcet-Houben M, Ballester A-R, Abbruscato P, González-Candelas L, Gabaldón T, Spadaro D. 2016. Genome sequencing and secondary metabolism of the postharvest pathogen Penicillium griseofulvum. BMC Genomics 17:19–19.
18. Balraj J, Jairaman K, Kalieswaran V, Jayaraman A. 2018. Bioprospecting lovastatin production from a novel producer Cunninghamella blakesleeana. 3 Biotech, 8(8):359–359.
19. Meneghetti F, Ferraboschi P, Grisenti P, Reza Elahi S, Mori M, Ciceri S. 2020. Crystallographic and NMR Investigation of Ergometrine and Methylergometrine, Two Alkaloids from Claviceps purpurea. Mol Basel Switz 25(2):331.
20. Brakhage AA. 2013. Regulation of fungal secondary metabolism. Nat Rev Microbiol 11(1):21–32.
21. Caboche S, Leclère V, Pupin M, Kucherov G, Jacques P. 2010. Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity. J Bacteriol, 192(19):5143–5150.
22. Parent A, Guillot A, Benjdia A, Chartier G, Leprince J, Berteau O. 2016. The B12-Radical SAM Enzyme PoyC Catalyzes Valine Cβ-Methylation during Polytheonamide Biosynthesis. J Am Chem Soc 138(48):15515–15518.
23. Curtis RW. 1958. Root Curvatures Induced by Culture Filtrates of Aspergillus niger. Science 128(3325):661.
24. Takeuchi S, Senn M, Curtis RW, McLafferty FW. 1967. Chemical studies on Malformin—V.: Malformin B1 and B2. Phytochemistry 6(2):287–292.
25. Kim K-W, Sugawara F, Yoshida S, Murofushi N, Takahashi N, Curtis W R. 1993. Structure of Malformin B, a Phytotoxic Metabolite Produced by Aspergillus niger. Biosci Biotechnol Biochem 57(5):787–791.
26. Suda S, Curtis RW. 1966. Antibiotic properties of malformin. Appl Microbiol 14(3):475–476.
27. Curtis RW. 1976. A Mung Bean Assay for Malformin-induced Growth Stimulation. Plant Physiol 57(3):365–368.
28. Schuster E, Dunn-Coleman N, Frisvad J, Dijck P. 2002. On the safety of Aspergillus niger - A review. Appl Microbiol Biotechnol 59(4-5):426–35.
29. Iriuchijima S, Curtis RW. 1969. Malformins from Aspergillus ficuum, A. awam ori and A. phoenicis. Phytochemistry 8(8):1397–1399.
30. Zhan J, Gunaherath GMKB, Wijeratne EMK, Gunatilaka AAL. 2007. Asperpyrone D and other metabolites of the plant-associated fungal strain Aspergillus tubingensis. Phytochemistry 68(3):368–372.
31. Quintanilha-Peixoto G, Torres RO, Reis IMA, Oliveira TAS de, Bortolini DE, Duarte EAA, Ariston de Carvalho Azevedo V, Brenig B, Aguiar ERGR, Soares ACF, Góes-Neto A, Branco A. 2019. Calm Before the Storm: A Glimpse into the Secondary Metabolism of Aspergillus welwitschiae, the Etiologic Agent of the Sisal Bole Rot. Toxins 11(11):631.
32. Marahiel MA, Essen L -O. 2009. Chapter 13 Nonribosomal Peptide Synthetases: Mechanistic and Structural Aspects of Essential Domains, p. 337–351. *In* Methods in Enzymology. Academic Press.
33. Cosmina P, Rodriguez F, de Ferra F, Grandi G, Perego M, Venema G, van Sinderen D. 1993. Sequence and analysis of the genetic locus responsible for surfactin synthesis in Bacillus subtilis. Mol Microbiol 8(5):821–831.

34. Team:Heidelberg. (2020, April 17).
Non-Ribosomal Peptide Synthesis. Get to Know the Theory. iGEM.
2013.igem.org/wiki/index.php?title=Team:Heidelberg/NRPS&oldid=353123.
35. Ali H, Ries MI, Lankhorst PP, van der Hoeven RAM, Schouten OL, Noga M, Hankemeier T, van Peij NNME, Bovenberg RAL, Vreeken RJ, Driessen AJM. 2014. A non-canonical NRPS is involved in the synthesis of fungisporin and related hydrophobic cyclic tetrapeptides in Penicillium chrysogenum. PloS One 9:e98212–e98212.
36. Röttig M, Medema MH, Blin K, Weber T, Rausch C, Kohlbacher O. 2011. NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res, 39:W362–W367.
37. Bozhüyük KAJ, Fleischhacker F, Linck A, Wesche F, Tietze A, Niesert C-P, Bode HB. 2018. De novo design and engineering of non-ribosomal peptide synthetases. Nat Chem 10:275–281.
38. Kjærbølling I, Vesth TC, Frisvad JC, Nybo JL, Theobald S, Kuo A, Bowyer P, Matsuda Y, Mondo S, Lyhne EK, Kogle ME, Clum A, Lipzen A, Salamov A, Ngan CY, Daum C, Chiniquy J, Barry K, LaButti K, Haridas S, Simmons BA, Magnuson JK, Mortensen UH, Larsen TO, Grigoriev IV, Baker SE, Andersen MR. 2018. Linking secondary metabolites to gene clusters through genome sequencing of six diverse Aspergillus species. Proc Natl Acad Sci 115:E753.
39. McLean TC, Wilkinson B, Hutchings MI, Devine R. 2019. Dissolution of the Disparate: Co-ordinate Regulation in Antibiotic Biosynthesis. Antibiotics 8(2):83.
40. Shelest E. 2008. Transcription factors in fungi. FEMS Microbiol Lett 286(2):145–151.
41. Oakley CE, Ahuja M, Sun W-W, Entwistle R, Akashi T, Yaegashi J, Guo C-J, Cerqueira GC, Russo Wortman J, Wang CCC, Chiang Y-M, Oakley BR. 2017. Discovery of McrA, a master regulator of Aspergillus secondary metabolism. Mol Microbiol, 103(2):347–365.
42. Perrin RM, Fedorova ND, Bok JW, Cramer RA, Wortman JR, Kim HS, Nierman WC, Keller NP. 2007. Transcriptional regulation of chemical diversity in Aspergillus fumigatus by LaeA. PLoS Pathog 3:e50–e50.
43. Bayram Ö, Krappmann S, Ni M, Bok JW, Helmstaedt K, Valerius O, Braus-Stromeyer S, Kwon N-J, Keller NP, Yu J-H, Braus GH. 2008. VelB/VeA/LaeA Complex Coordinates Light Signal with Fungal Development and Secondary Metabolism. Science 320(5882):1504.
44. Patananan AN, Palmer JM, Garvey GS, Keller NP, Clarke SG. 2013. A novel automethylation reaction in the Aspergillus nidulans LaeA protein generates S-methylmethionine. J Biol Chem, 288(20):14032–14045.
45. Barkess G, West AG. 2012. Chromatin insulator elements: establishing barriers to set heterochromatin boundaries. Epigenomics 4(1):67–80.
46. Pfannenstiel BT, Keller NP. 2019. On top of biosynthetic gene clusters: How epigenetic machinery influences secondary metabolism in fungi. Biotechnol Adv 37(6):107345.
47. Collemare J, Seidl MF. 2019. Chromatin-dependent regulation of secondary metabolite biosynthesis in fungi: is the picture complete? FEMS Microbiol Rev 43(6):591-607.
48. Bok JW, Chiang Y-M, Szewczyk E, Reyes-Dominguez Y, Davidson AD, Sanchez JF, Lo H-C, Watanabe K, Strauss J, Oakley BR, Wang CCC, Keller NP. 2009. Chromatin-level regulation of biosynthetic gene clusters. Nat Chem Biol 5(7):462–464.

49. Macheleidt J, Mattern DJ, Fischer J, Netzker T, Weber J, Schroeckh V, Valiante V, Brakhage AA. 2016. Regulation and Role of Fungal Secondary Metabolites. Annu Rev Genet 50:371–392.
50. Boruta T. 2018. Uncovering the repertoire of fungal secondary metabolites: From Fleming's laboratory to the International Space Station. Bioengineered 9(1):12–16.
51. Pan J, Bhardwaj M, Nagabhyru P, Grossman RB, Schardl CL. 2014. Enzymes from Fungal and Plant Origin Required for Chemical Diversification of Insecticidal Loline Alkaloids in Grass-Epichloë Symbiota. PLOS ONE 9:e115590.
52. Abdel-Azeem A, Salem F, Abdel Azeem M, Nafady N, Mohesien M, Soliman EA. 2016. Biodiversity of the Genus Aspergillus in Different Habitats, p. 3–28. *In* .
53. Tsang A, Bellemare A, Darmond C, Bakhuis J. 2015. Genetic and Genomic Manipulations in Aspergillus niger, p. 225–243. *In* van den Berg, MA, Maruthachalam, K (eds.), Genetic Transformation Systems in Fungi, Volume 2. Springer International Publishing, Cham.
54. Aguilar-Pontes MV, Brandl J, McDonnell E, Strasser K, Nguyen TTM, Riley R, Mondo S, Salamov A, Nybo JL, Vesth TC, Grigoriev IV, Andersen MR, Tsang A, de Vries RP. 2018. The gold-standard genome of Aspergillus niger NRRL 3 enables a detailed view of the diversity of sugar catabolism in fungi. Stud Mycol 91:61–78.
55. Fowler T, Berka RM, Ward M. 1990. Regulation of the glaA gene of Aspergillus niger. Curr Genet 18:537–545.
56. Zhu X, Wang M, Qiu R, Liu L, Dong Z, Tang G. 2004. The synergetic effects of two CCAAT boxes in Aspergillus niger glaA gene promoter on activation of PglaA transcription. Sci China Ser C Life Sci Chin Acad Sci 47(2):139–47.
57. Zhu X, Qiu R, Liu L, Tang G. 2004. Binding of trans-acting protein AngCP to the CCAAT-containing motifs in Aspergillus niger glaA promoter. Prog Nat Sci 14:338–343.
58. Tanaka A, Kamei K, Tanoue S, Papagiannopoulos P, Steidl S, Brakhage AA, Davis MA, Hynes MJ, Kato M, Kobayashi T, Tsukagoshi N. 2001. AoHapB, AoHapC and AoHapE, subunits of the Aspergillus oryzae CCAAT-binding complex, are functionally interchangeable with the corresponding subunits in Aspergillus nidulans. Curr Genet 39:175–182.
59. Palys S. 2017. A bioinformatics characterization of secondary metabolism and alkyl citric acid pathway reconstruction in Aspergillus niger NRRL3. Concordia University.
60. Inglis DO, Binkley J, Skrzypek MS, Arnaud MB, Cerqueira GC, Shah P, Wymore F, Wortman JR, Sherlock G. 2013. Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of Aspergillus nidulans, A. fumigatus, A. niger and A. oryzae. BMC Microbiol 13:91.
61. Zabala AO, Xu W, Chooi Y-H, Tang Y. 2012. Characterization of a silent azaphilone gene cluster from Aspergillus niger ATCC 1015 reveals a hydroxylation-mediated pyran-ring formation. Chem Biol 19(8):1049–1059.
62. Aerts D, Hauer EE, Ohm RA, Arentshorst M, Teertstra WR, Phippen C, Ram AFJ, Frisvad JC, Wösten HAB. 2018. The FlbA-regulated predicted transcription factor Fum21 of Aspergillus niger is involved in fumonisin production. Antonie Van Leeuwenhoek, 111(3):311–322.
63. Holm DK, Petersen LM, Klitgaard A, Knudsen PB, Jarczynska ZD, Nielsen KF, Gotfredsen CH, Larsen TO, Mortensen UH. 2014. Molecular and Chemical Characterization of the Biosynthesis of the 6-MSA-Derived Meroterpenoid Yanuthone D in Aspergillus niger. Chem Biol 21(4):519–529.

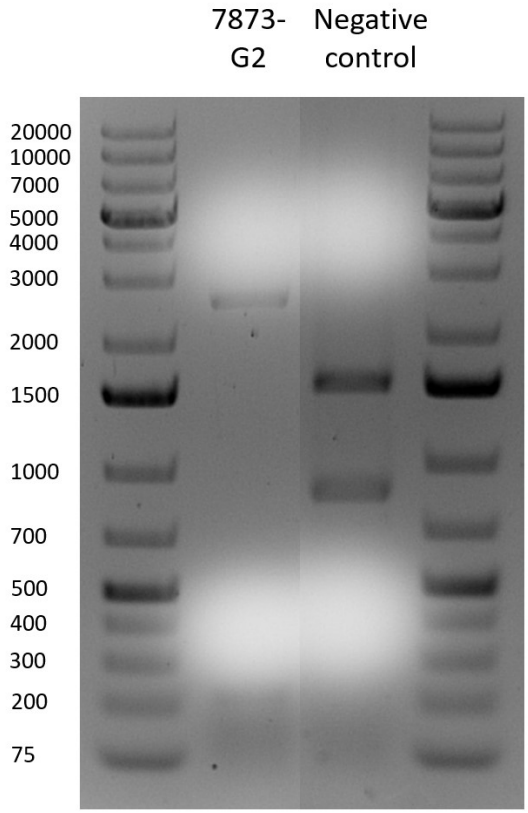
64. Li Y, Chooi Y-H, Sheng Y, Valentine JS, Tang Y. 2011. Comparative characterization of fungal anthracenone and naphthacenedione biosynthetic pathways reveals an α-hydroxylation-dependent Claisen-like cyclization catalyzed by a dimanganese thioesterase. J Am Chem Soc, 133(39):15773–15785.
65. Awakawa T, Yang X-L, Wakimoto T, Abe I. 2013. Pyranonigrin E: A PKS-NRPS Hybrid Metabolite from Aspergillus niger Identified by Genome Mining. ChemBioChem 14(16):2095–2099.
66. Yang X-L, Awakawa T, Wakimoto T, Abe I. 2014. Three Acyltetronic Acid Derivatives: Noncanonical Cryptic Polyketides from Aspergillus niger Identified by Genome Mining. ChemBioChem 15(11):1578–1583.
67. Palys S, Pham TTM, Tsang A. 2019. Biosynthesis of alkylcitric acids in Aspergillus niger involves both co-localized and unlinked genes. bioRxiv 714071.
68. Meyer V, Arentshorst M, El-Ghezal A, Drews A-C, Kooistra R, van den Hondel CAMJJ, Ram AFJ. 2007. Highly efficient gene targeting in the Aspergillus niger kusA mutant. J Biotechnol 128(4):770–775.
69. Ho C, Lam C, Chan M, Cheung R, Law L, Lit L, Ng K, Suen M, Tai H. 2003. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. Clin Biochem Rev 24(1):3–12.
70. Hur GH, Vickery CR, Burkart MD. 2012. Explorations of Catalytic Domains in Non-Ribosomal Peptide Synthetase Enzymology. Nat Prod Rep 29(10):1074–1098.
71. Elgubbi H, Maltin A, Alzaridy A. 2015. Modified Ninhydrin Reagent for the Detection of Amino Acids on TLC Paper. EC Nutr 1:128–131.
72. Pang B, Chen Y, Gan F, Yan C, Jin L, Gin JW, Petzold CJ, Keasling JD. 2020. Investigation of Indigoidine Synthetase Reveals a Conserved Active-Site Base Residue of Nonribosomal Peptide Synthetase Oxidases. J Am Chem Soc 142(25):10931–10935.
73. Tong S, Chern S-WW, Li Y, Pallansch MA, Anderson LJ. 2008. Sensitive and Broadly Reactive Reverse Transcription-PCR Assays To Detect Novel Paramyxoviruses. J Clin Microbiol 46(8):2652–2658.
74. Ogasawara Y, Dairi T. 2018. Peptide Epimerization Machineries Found in Microorganisms. Front Microbiol 9:156.
75. von der Haar T, Leadsham JE, Sauvadet A, Tarrant D, Adam IS, Saromi K, Laun P, Rinnerthaler M, Breitenbach-Koller H, Breitenbach M, Tuite MF, Gourlay CW. 2017. The control of translational accuracy is a determinant of healthy ageing in yeast. Open Biol 7.
76. van Dijk JWA, Guo C-J, Wang CCC. 2016. Engineering Fungal Nonribosomal Peptide Synthetase-like Enzymes by Heterologous Expression and Domain Swapping. Org Lett 18(24):6236–6239.
77. Sanchez JF, Somoza AD, Keller NP, Wang CCC. 2012. Advances in Aspergillus secondary metabolite research in the post-genomic era. Nat Prod Rep, 29(3):351–371.
78. Xia J, Bjorndahl TC, Tang P, Wishart DS. 2008. MetaboMiner – semi-automated identification of metabolites from 2D NMR spectra of complex biofluids. BMC Bioinformatics 9:507.
79. Kiontke A, Oliveira-Birkmeier A, Opitz A, Birkemeyer C. 2016. Electrospray Ionization Efficiency Is Dependent on Different Molecular Descriptors with Respect to Solvent pH and Instrumental Configuration. PLOS ONE 11:e0167502.
80. Feige MJ. 2018. Oxidative Folding of Proteins: Basic Principles, Cellular Regulation and Engineering. Royal Society of Chemistry.

81. Rendsvig JKH, Workman CT, Hoof JB. 2019. Bidirectional histone-gene promoters in Aspergillus: characterization and application for multi-gene expression. Fungal Biol Biotechnol 6:24–24.
82. Gade P, Kalvakolanu DV. 2012. Chromatin Immunoprecipitation Assay as a Tool for Analyzing Transcription Factor Activity. Methods Mol Biol Clifton NJ 809:85–104.
83. Jomori T, Hara Y, Sasaoka M, Harada K, Setiawan A, Hirata K, Kimishima A, Arai M. 2020. Mycobacterium smegmatis alters the production of secondary metabolites by marine-derived Aspergillus niger. J Nat Med 74(1):76–82.
84. Takahashi H, Morioka R, Ito R, Oshima T, Amin A, Ogasawara N, Kanaya S. 2011. Dynamics of Time-Lagged Gene-to-Metabolite Networks of Escherichia coli Elucidated by Integrative Omics Approach. Omics J Integr Biol 15(1-2):15–23.
85. Nakabayashi R, Mori T, Nishizawa T, Saito K. 2017. Temporal lag between gene expression and metabolite accumulation in flavonol biosynthesis of Arabidopsis roots. Phytochem Lett 22:44–48.
86. Steyn PS, Vleggaar R, Simpson TJ. 1984. Stable isotope labelling studies on the biosynthesis of asticolorin C by Aspergillus multicolor. Evidence for a symmetrical intermediate. J Chem Soc Chem Commun 765–767.
87. Akashi T, Oakley CE, Oakley BR. 2020. How McrA regulates secondary metabolism.
88. Grau MF, Entwistle R, Oakley CE, Wang CCC, Oakley BR. 2019. Overexpression of an LaeA-like Methyltransferase Upregulates Secondary Metabolite Production in Aspergillus nidulans. ACS Chem Biol 14(7):1643–1651.
89. Song L, Ouedraogo J-P, Kolbusz M, Nguyen TTM, Tsang A. 2018. Efficient genome editing using tRNA promoter-driven CRISPR/Cas9 gRNA in Aspergillus niger. PloS One 13(8):e0202868.

Supplemental

Quality control of *NRRL_07873* strains

A forward primer starting 45 base pairs upstream the start codon and a reverse primer ending 92 base pairs downstream of the stop codon was used to amplify the coding region of any gene with a GlaA promoter and terminator. *NRRL3_07873* in particular is 97 bp longer than the GlaA gene making it difficult to resolve the difference on an agarose gel. While the *NRRL3_07873* band did look larger than the GlaA band on an electrophoresis gel, a restriction digest was used to resolve some of the ambiguity. A restriction enzyme that cuts a restriction site found within the GlaA gene but not the *NRRL3_07873* gene, BglII, was used. The *NRRL3_07873* gene lacks a unique cut site that could be used to confirm its presence with a positive result. The 2269 base pair *NRRL_07873* gene is expected to produce a 2406 base pair band while the GlaA gene of the negative control is expected to be cut and produce a 1476 base pair band and an 833 base pair band. The amplicons for this PCR can be seen run on an electrophoresis gel in supplemental figure 1.

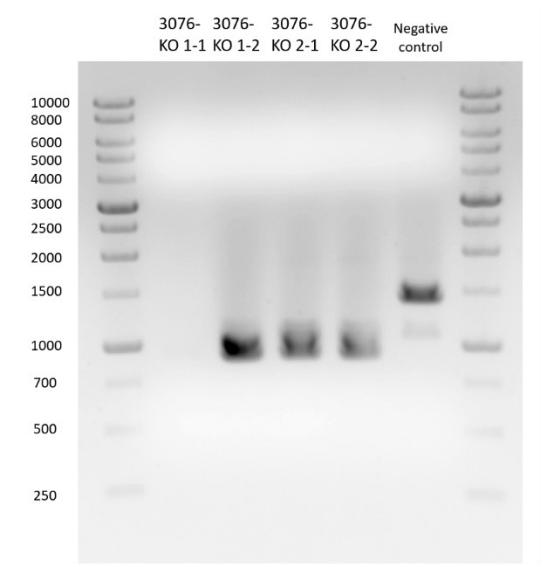


Supplemental figure 1. Amplification of glucoamylase locus in *NRRL3_07873* targeted integration strain. Electrophoresis agarose gel of amplicons amplified with TFChk_45_Fw and TFChk_92_Rv primers and digested with BglII restriction enzyme, run at 100 volts. 7873-G2 denotes the second colony taken from a transformation where the *NRRL3_07873* gene was inserted into the glucoamylase locus.

McrA depletion strain quality control

A forward primer starting 749 base pairs upstream of the start codon and a reverse primer ending 675 base pairs downstream of the start codon were used to check for the deleted region of the McrA gene as a means of quality control. Lack of deletion is expected to produce an amplicon of 1424 base pairs while a deletion of the 485 base pair region is expected to produce an amplicon of 939 base pairs.

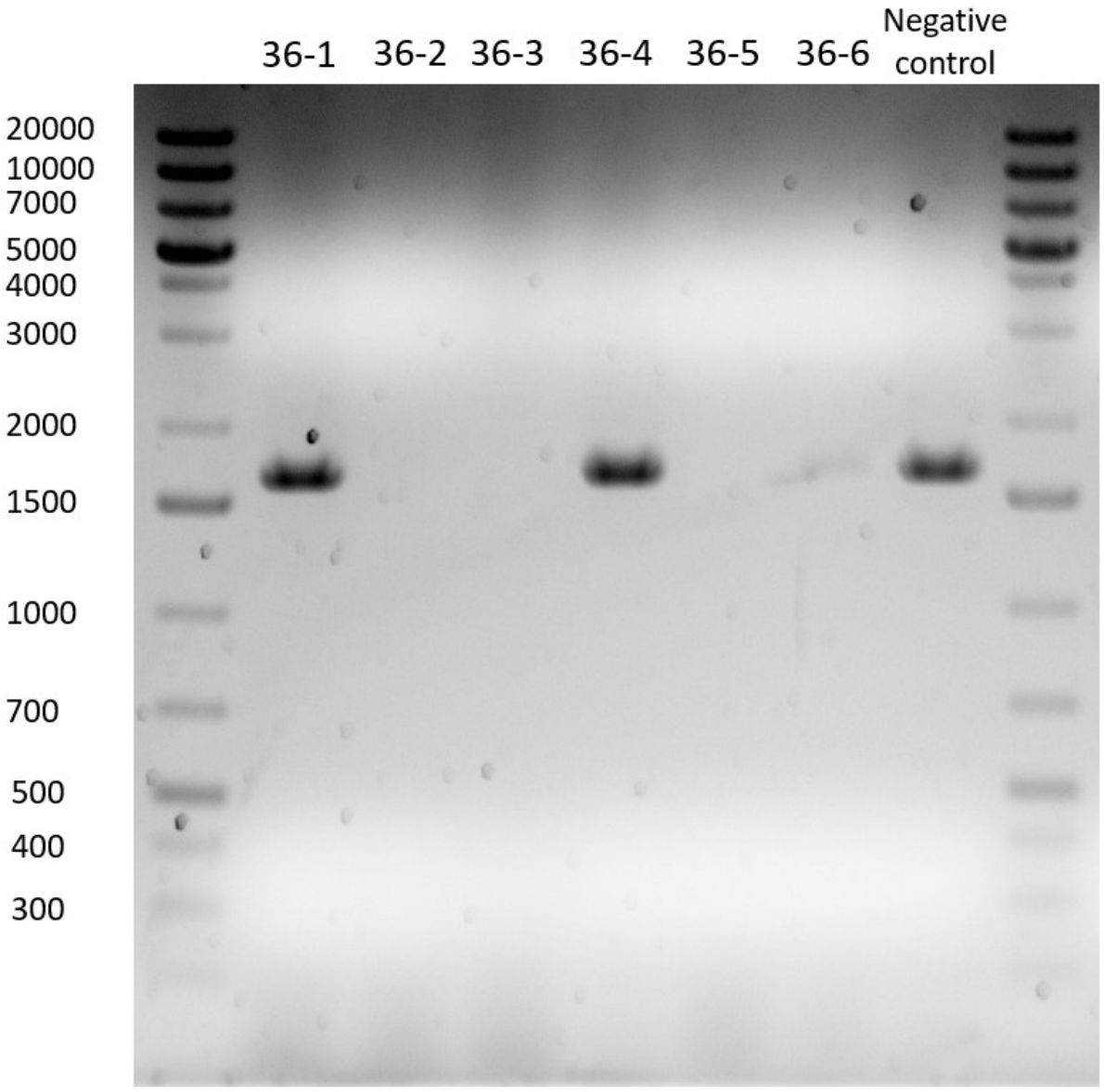
Figure 10 shows the amplicons amplified from genomic DNA from two McrA depletion strains run in duplicate and the parental strain.



Supplemental figure 2. Electrophoresis agarose gel of amplicons amplified with 3076KOChkFw and 3076KOChkRv primers, run at 100 volts.

Quality control *NRRL3_00036* deletion strains

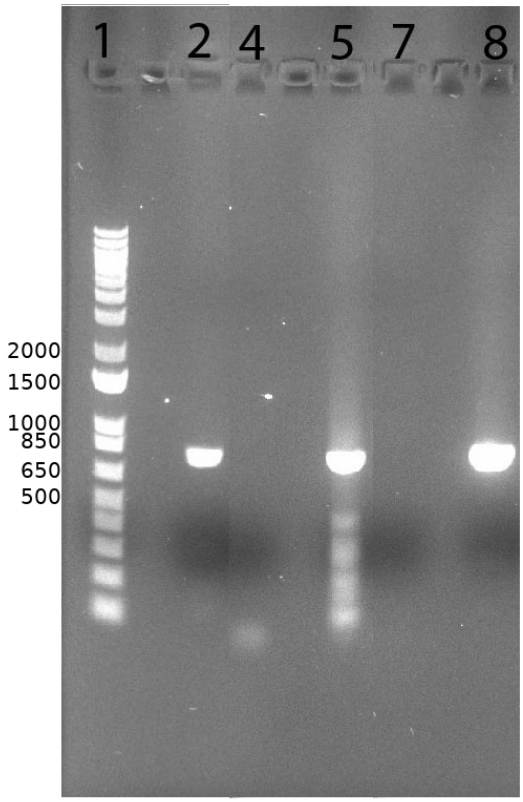
A forward primer starting 920 base pairs before the start codon and a reverse primer ending 743 base pairs downstream of the start codon were used as flanking primers to check for deletion of the *NRRL3_00036* gene. Amplicons for this PCR can be seen run on an electrophoresis gel in figure 15.



Supplemental figure 3. Amplification of the *NRRL3_00036* locus in *NRRL3_00036* deletion strain with flanking primers. Electrophoresis agarose gel of amplicons amplified with 36KO_Chk_Fw and 36KO_Chk_Rv primers, run at 100 volts

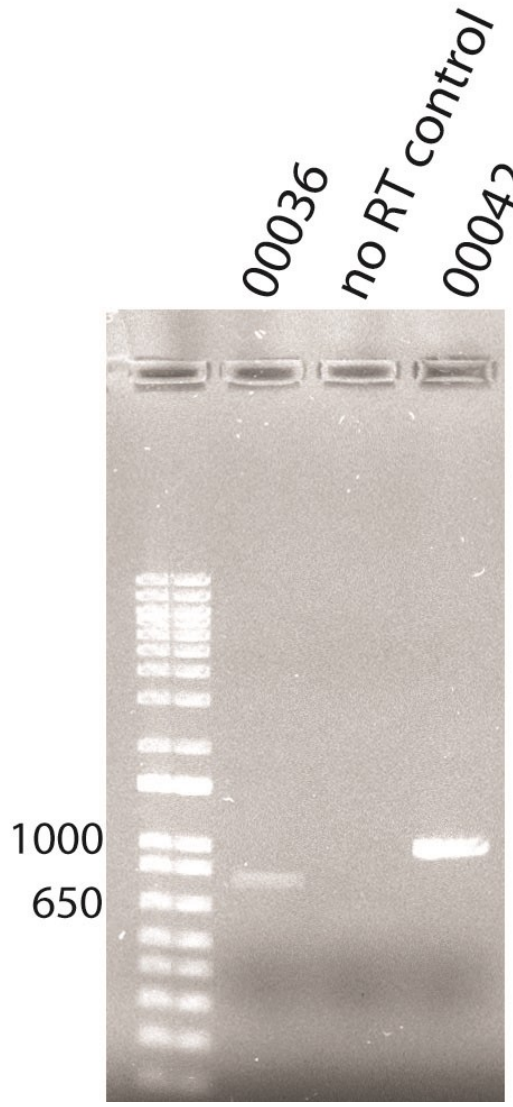
NRR3_00036 cluster gene expression by reverse transcription PCR

Expression of the *NRRL3_00042* and *NRRL3_00036* genes in the *NRRL3_00042* overexpression strain was tested by reverse transcription PCR (supplemental figure 4).



Supplemental figure 4. Reverse transcription PCR of the *NRRL3_00036* gene in the *NRRL3_00042* overexpression strain and parental strain

Lane 1: Ladder, Lane 2: cDNA sourced from parental strain at 5-days - amplification using *NRRL3_00036* primers, Lane 4: no reverse transcriptase control from parental strain at 5-days - amplification using *NRRL3_00036* and *NRRL3_00042* primers (multiplex), Lane 5: cDNA sourced from parental strain at 7-days - amplification using *NRRL3_00036* primers, Lane 7: no reverse transcriptase control from parental strain at 7-days - amplification using *NRRL3_00036* and *NRRL3_00042* primers (multiplex), Lane 8: cDNA sourced from *NRRL3_00042* overexpression strain at 7-days - amplification using *NRRL3_00036* primers.



Supplemental figure 5. Reverse transcription PCR of the *NRRL3_00042* overexpression strain. Results are from RNA extracted at 7 says of growth

Supplemental table 1. Primer and repair template table:

TFChk_45_Fw	AACTGAGAGCCTGAGCTTC
TFChk_92_Rv	TCACTACTATTATGCACACCC
NRRL3_03076 gRNA	GCAACATACAGGATACTGAG
NRRL3_03076 repair template	GCCTTCCGTACATACTTTACGTACTTACATAGTATCCTGTATGT TGCTTGCTCTTTTGCT
3076KOChkFw	CATCCTTCCCTGACCCTTGC

3076KOChkRv	GACAGAATCGACGCGAGAACG
NRRL3_00036 gRNA	GTACACCCGCAACTTTACCT
NRRL3_00036 60 base pair repair template	CCGCAGGCACATCTCAGCTCGCATGTCGACCATCAAACCGGACCATCCCCAATGCAGTGT
NRRL3_00036 90 base pair repair template	TAGCGGGTCTCAATTCCGCAGGCACATCTCAGCTCGCATGTCGACCATCAAACCGGACCATCCCCAATGCAGTGTCTAAGCAACATCCCG
36KO_Chk_Fw	GTGGATGTCCAAGCAACCAC
36KO_Chk_Rv	GTGATAGCACGACCGTTGATG