Mining for Metabolites: A study of *Alternaria* species'

Nicolas Villeneuve

A thesis submitted to the Faculty of Science
in partial fulfillment of the requirements for the degree of

Master of Science

in Chemistry

Specialization in Metabolomics

Department of Chemistry and Biomolecular Sciences
Faculty of Science
University of Ottawa

Abstract

Over the past several years, fungal species' (*spp*.) with the capability to infect humans have caused an estimated 1.6 million deaths annually, and are threatening bats, amphibians, and reptiles with extinction (Frick et al., 2010; Brown et al., 2012; Casadevall, 2017). Additionally, contamination of crops by fungal pathogens causes a loss of a third of all of the crops grown annually, thereby causing billions of dollars in economic losses on an annual basis (Fisher et al., 2012; Almeida et al., 2019). A common crop contaminant is the fungal genera comprised of many fungal pathogens, *Alternaria*.

The overall objective of the present thesis was to use untargeted LC-MS metabolomics to further our understanding of their pathogenicity. Specifically, a robust profile for 60 strains of uncharacterized *Alternaria spp*. was generated which, when mined, revealed the presence of several unique secondary metabolites (SMs). Subsequent isolation, purification, and structural elucidation of those SMs revealed them to be the family of curvularins. Mining the genome of the isolates producing those curvularins revealed the biosynthetic gene cluster (BGC) necessary for the production of the curvularins. Generating sequence similarity networks (SSNs) allowed for the investigation of the similarity of the BGC necessary for the synthesis of the curvularins, across several pathogenic fungal *spp*.

Acknowledgments

I would like to express my deepest appreciation to my thesis supervisors, Dr. Dave Overy and Dr. Chris Boddy, for all of the help and guidance they have provided me over the past two years. Additionally, I am extremely grateful to my thesis advisory committee, who has helped guide me through these two years of trying times. I would like to thank Amanda Sproule for all of the excellent data she obtained for me, Quinn Eggertson for all of the in-lab help, and Dr. Jeremy Dettman for all of the whole genome sequences he obtained and allowed me to utilize for this thesis. Finally, I would like to thank everyone else in the Metabolomics laboratories of the Central Experimental Farm, who were extremely kind and helpful.

Table of Contents

List of Figures

List of Tables

Abbreviations

ACP	Acyl-carrier protein
ACs	Accessory chromosomes
AT	Acyltransferase
BGCs	Biosynthetic gene cluster
DH	Dehydratase
DHC	Dehydrocurvularin
ER	Enoylreductase
ESI	Electrospray ionization
EtOAc	Ethyl acetate
EtOH	Ethanol
HGT	Horizontal gene transfer
HSTs	Host-specific-toxins
KR	Ketoreductase
KR_c	KR catalytic subdomain
KR_s	KR structural subdomain
KS	Ketosynthase
LC-MS	Liquid chromatography – Mass Spectrometry
Mass-to-charge	*m/z*
MAT	Malonyl-CoA:ACP acytransferase
MeOH	Methanol
MS	Mass spectrometry
nHSTs	Non-host-specific-toxins
NMR	Nuclear magnetic resonance
PKR	Pseudo ketoreductase
PMT	Pseudo C-methyltransferase
RT	Retention time
SMs	Secondary metabolites
Spp.	species'
SSN	Sequence similarity network
TLC	Thin layer chromatography
UPLC	Ultra performance liquid chromatography

Chapter 1

1 Introduction

1.1 Overview

The genome is the complete set of genetic information in an organism, containing all of the information required for the organism to function (Nature Education, 2014). In all living organisms, the genome is comprised of DNA which form packaged structures called chromosomes (Cooper, 2000). The genome of prokaryotes (single-celled organisms) is generally packaged into a single circular chromosome structure (Cooper, 2000). The genome of eukaryotes (multicellular organisms), by contrast, is packaged into several chromosomes, of varying sizes and shapes (Cooper, 2000).

The information encoded in the genome is transcribed into individual transportable cassettes through a process called transcription, whereby, the DNA is converted to messenger ribonucleic acid (mRNA) (Liu, 2019). Transcription is the process where transcription machinery reads the information encoded in the DNA and produces the exact complement of that information as mRNA (Lodish et al., 2000). The sequence of mRNA is then made into proteins (i.e. enzymes) through a process called translation (Lengyel, 1969). Proteins catalyze the metabolic reactions (i.e. both the anabolic and catabolic) performed by an organism - the reactions necessary to assemble small molecules into large ones, or to break down large molecules to small ones (Bartee et al., 2017). The products of metabolic processes are generally termed metabolites and,

as such, the collection of metabolites of a given organism is generally referred to as the metabolome (Patti et al., 2012).

Primary metabolites are the molecules involved in primary metabolism, which is: the biochemical processes essential to basic cellular function (i.e. molecules involved in energy generation, growth, and reproduction) (Erb and Kliebenstein, 2020). Secondary metabolites (SMs) are metabolites which do not serve a function essential to life, but rather, they confer an increased fitness to the producer (Erb and Kliebenstein, 2020). They are often the result of active selection for the biosynthetic processing of primary intermediates or metabolites, catalyzed by unique enzymes for accomplishing such tasks (Vining, 1992). One could say primary metabolites are the molecules necessary for an organism to survive, while SMs are necessary for an organism to thrive.

Transcriptomics is the analysis of the complete complement of mRNA molecules in a cell, tissue or organ, denoting the sequence and quantity of mRNA molecules transcribed in that system (the transcriptome). Though the transcriptome is subject to the changing physiological, pathological or developmental state of the system, the transcriptome has a direct relationship with the genome and could therefore be studied using similar techniques to genomics. Proteomics is the analysis of the complete complement of protein molecules in a system (the proteome) (Kell and Oliver, 2016).

Metabolomics, therefore, is the application of data science to the complete set of metabolites produced in a system (the metabolome). When compared to the other levels of 'omics it

represents the most complex to decipher and link to the genome as many genes may determine the synthesis and turnover of a single metabolite and they may all be subject to both extrinsic and intrinsic factors (Lowe et al., 2017). The metabolome, however, poses the benefit of being directly linked to function and allows the dynamic and sensitive measure of the phenotype at the molecular level (Johnson et al., 2016; Rinschen et al., 2019).

1.1.1 Approaches to Metabolomics Experiments

In metabolomic studies, the approach to experimental design may be either targeted or untargeted. When a targeted approach is taken, the identity and importance of the specific set of metabolites in question has already been hypothesized/elucidated prior to study (Patti et al., 2012; Zamboni et al., 2015). As such, one may optimize their sample preparation and analytical workflow accordingly to achieve increased sensitivity and precision towards the detection/quantification of that/those set(s) of targeted metabolites (Patti et al., 2012; Zamboni et al., 2015). When the purpose of the experiment is not to identify/detect a specific set of metabolites, but rather, to identify/probe as many metabolites as possible, an untargeted approach is taken (Patti et al., 2012).

When an untargeted approach is taken, experimental design and conditions are often very different from a targeted approach. For example, for the study of the expression of fungal metabolism in culture, an untargeted approach would use multiple culture conditions/media formulations to promote maximal metabolite diversity versus the use of a specific condition/medium formulation to maximize the expression of a particular set of metabolites or to focus on a particular metabolic biosynthetic pathway. These approaches make metabolomics powerful in identifying silent-phenotypes, genes expressed under stress which otherwise usually

have no apparent influence on physical characteristics or behaviour and thereby go unnoticed in the other 'omic techniques (Griffin et al., 2002). Complemented with the other 'omic techniques, metabolomic studies can allow for the mechanistic elucidation of interesting natural products (Zhang et al., 2021).

1.1.2 Targeted Metabolomics

Targeted metabolomics is the measurement of defined groups of chemically-characterized and biochemically-annotated metabolites, typically focusing on one or few related pathways of interest (Dudley et al., 2010; Patti et al., 2012). Targeted metabolomics experiments are driven by a specific biochemical question or hypothesis that requires the investigation of a particular pathway (Patti et al., 2012). As such, they often employ the use of native and isotopically-labelled internal standards of the pathway in question (Roberts et al., 2012). This allows the experimenter to reduce the frequency of false positives, to help account for matrix-induced ionization effects which affect analytical precision and metabolite resolution, and to reduce the bias caused by the dominance of high-abundance molecules in the analyses (Patti et al., 2012; Roberts et al., 2012). Additionally, since targeted metabolomics often involves the clear definition of at least one molecule of interest, data processing steps are focused on clearly retrieving the molecule(s) of interest (Roberts et al., 2012; Ribbenstedt et al., 2018). The predominant drawback of targeted metabolomic studies is that when only focusing on a specific set of metabolites, the thousands of remaining metabolites of the metabolome are ignored (Ribbenstedt et al., 2018).

1.1.3 Equipment Used in Metabolomics

Regardless the taken approach, sophisticated tools of analytical chemistry (liquid chromatography (LC), mass spectrometry (MS), and nuclear magnetic resonance (NMR) spectroscopy) are required to study the metabolome. The metabolomics approach used in this thesis, focused upon LC-MS based metabolomics. A common pipeline for metabolomics experiments, begins with the initial data acquisition phase. For example, when using a LC-MS platform, the initial data is usually acquired using MS, which when coupled to an LC, allows for a clear separation of the metabolites that comprise the sample of interest prior to detection. Each of the metabolite features detected using MS correspond to the detection of an ion with a unique/specific mass-to-charge (*m/z*) ratio and retention time (Patti et al., 2012).

A downside of LC-MS for initial data acquisition, when using a soft ionization technique such as electrospray ionization (ESI), is the layers of complexity brought on by the formation of adducts, neutral ion losses, and fragmentation that occur during ionization, that can be further confounded by the co-elution of different molecules. Additionally, the results of an LC-MS scan are not directly representative of the quantity of a given molecule in a sample. As such, LC-MS is generally the first step in data acquisition as it yields a good representation of the composition of a given sample (Jonsson et al., 2005; Barton et al., 2008; Nordström et al., 2008; Büscher et al., 2009; Patti et al., 2012).

The *m/z* ratio, retention time, and signal intensity are the defining characteristics of raw metabolite features and (to some extent) they are a factor/function of the MS ionization event. Several different ionization methods have been probed in the context of metabolomics and they

all discriminate differently and uniquely against certain analyte physical-chemical properties (Nordström et al., 2008). In the context of the present study, the initial data was acquired using an ESI source, which was preceded by an Ultra Performance Liquid Chromatography (UPLC) device.

1.1.4 Metabolomics pipeline/workflow

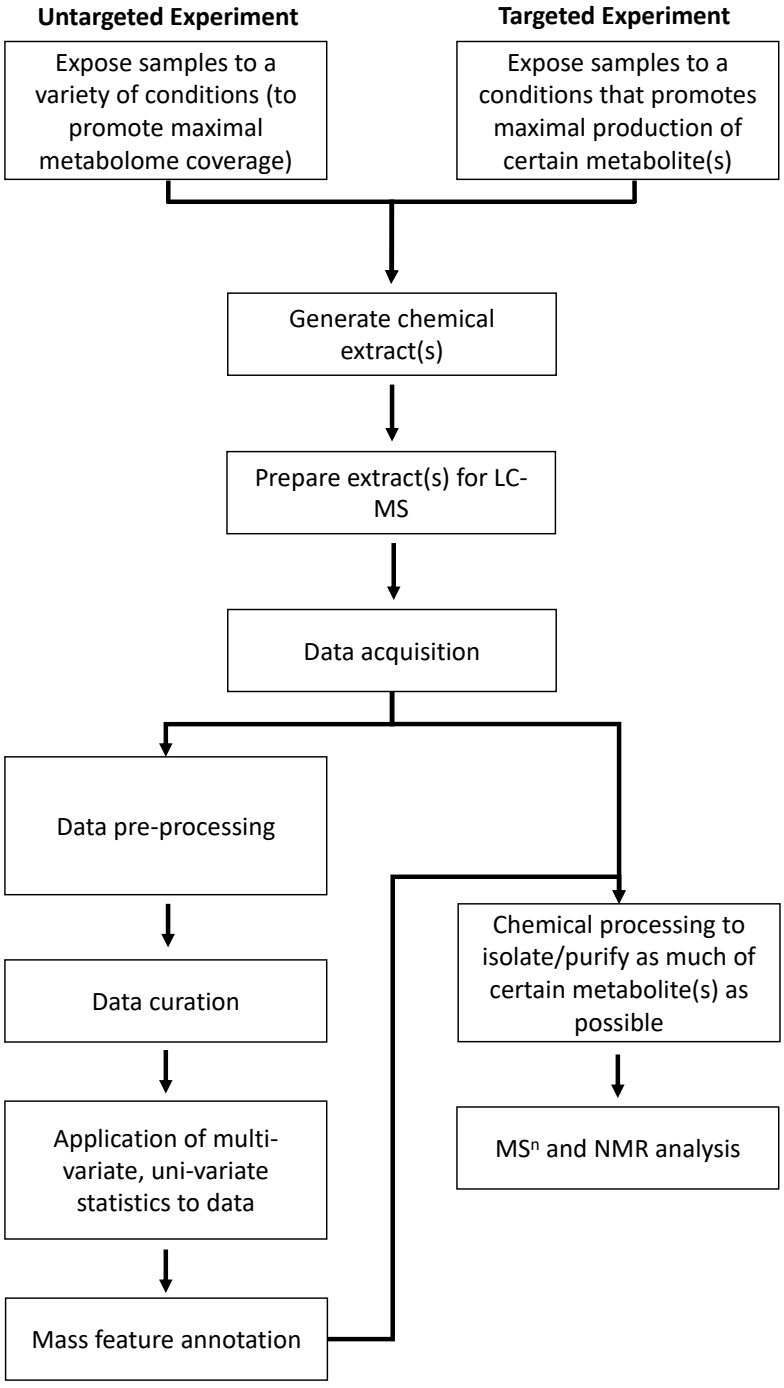

Figure 1.1. A flowchart depicting the general workflow of a targeted or untargeted metabolomics experiment.

1.1.5 Untargeted Metabolomics Pipeline

The workflow of LC-MS based untargeted metabolomics experiments may generally be broken down to several steps, illustrated by **Figure 1.1.** The first step in the process is to grow the selected samples under a variety of conditions, thereby, aiming to promote the maximal metabolome coverage being achieved. The next step is to generate chemical extracts from the samples. This is typically done using an extraction solvent like ethyl acetate (EtOAc), methanol (MeOH), ethanol (EtOH), or other solvents of a similar polarity range. The generation of chemical extract(s) is generally then followed by a filtration step followed by vacuum concentration of the extract(s), and their subsequent preparation for chromatographic separation and detection by LC-MS (data acquisition). The resulting data output is represented in a 3-dimensional plane, with the RT and *m/z* of the analytes denoting the metabolite feature and the relative intensity representing the associated ions of the metabolite(s) detected. The raw data then undergoes spectral pre-processing, curation, and subsequent statistical analyses. Both will be discussed in greater detail in the **Data Analysis** section of the current work. After data pre-processing, data curation, and the application of multi- and uni-variate statistics to the data, interesting mass features are generally annotated. Mass feature annotation often informs subsequent targeted metabolomic experimentation.

1.1.6 Targeted Metabolomics Pipeline

The workflow of a targeted metabolomics experiment is often informed by the results obtained by untargeted metabolomics experimentation, and a general workflow may be denoted by **Figure 1.1.** The difference between the workflow of a targeted and untargeted metabolomics experiment is in that the growth conditions utilized are to maximize the production of certain metabolites.

Alternatively, a targeted metabolomics approach can be utilized after annotation of the features in the untargeted metabolomics experiment, given that the metabolite of interest is present in sufficient quantity for chemical processing. The generation of chemical extracts, drying, and preparation for LC-MS are much the same between the two, though with the LC-MS information being utilized to confirm the presence of the metabolites of interest. If the metabolites of interest are detected in the bulk extract, chemical processing steps then follow. The chemical processing steps are generally with the intent of obtaining the maximal amount of the metabolites of interest as possible, thereby, allowing for the maximal possible amount to be utilized for structure elucidation, bioactivity determination, or etc.

1.1.7 Data Pre-processing

A typical raw MS data file resulting from an untargeted analysis can contain tens of thousands of data points (Bauer et al., 2011). As a result, no matter how high a resolution the utilized MS can boast, background noise is unavoidable. The main sources of noise are chemically-derived such as being from the solvent/matrix, sample contamination or carry-over, electrical noise, or noise from the instrument itself (Kwon et al., 2008; Gardner and Brodbelt, 2009; Bauer et al., 2011). As such, to set the stage for the accurate statistical analysis and biological interpretation of data obtained by LC-MS, it is first necessary that the data go through a workflow of pre-processing steps (**Figure 1.2**). These steps are often performed using MZMine, a well-supported open-source program for metabolomic pre-processing (which may be downloaded from: http://mzmine.github.io/).

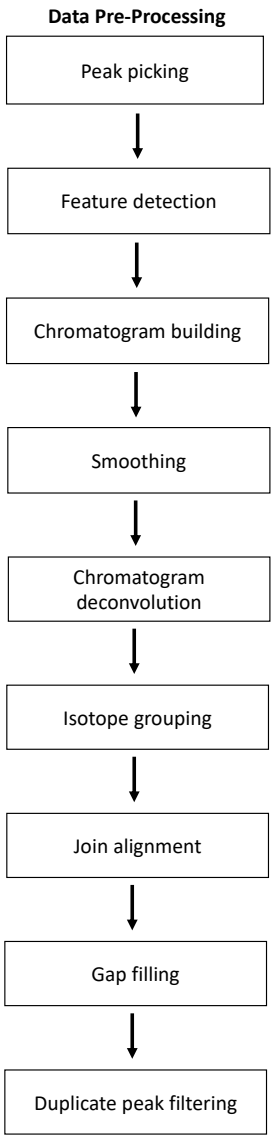

Figure 1.2. A flowchart depicting the general workflow of metabolomic data pre-processing

The first step in the data pre-processing pipeline is *peak picking*. *Peak picking* is the application of a peak picking algorithm to the spectral data, in order to separate the background noise from the metabolite peaks (Du et al., 2006; Lange et al., 2006; Bauer et al., 2011). Additionally, peak filtering can be used to exclude low abundant metabolite features and thus, minimize the number of false positives carried through to analysis (Du et al., 2006; Lange et al., 2006; Bauer et al.,

2011). Applying a *peak picking* algorithm to the raw data outputs the data in centroid format, thereby, allowing it to be imported into MZMine for the subsequent pre-processing steps.

The next step in the pre-processing pipeline is *feature detection*. *Feature detection* is where each MS spectrum is processed individually and converted to pairs of *m/z* and intensity values. In order to apply a *feature detection* algorithm, it is only required that a threshold *m/z* level be specified, under which, a peak would not be considered (Pluskal et al., 2010). In order to execute a *feature detection* algorithm in MZMine, a noise level threshold needs be specified.

The next step in the pre-processing pipeline is *chromatogram building*. *Chromatogram building* connects consecutive *m/z* values spanning multiple scans, over a certain duration of time, and bundles them as chromatogram objects. In order to apply the chromatogram building algorithm to the dataset, a *m/z* tolerance needs be specified (Pluskal et al., 2010). The next step in the process is optional, but highly recommended to reduce the occurrence of false positives in the dataset (Irizarry, 2019). An optional next step is the application of a *smoothing* algorithm to the dataset, whereby a filter width needs be specified (Pluskal et al., 2010).

The next step in the data pre-processing pipeline is *chromatogram deconvolution*. Applying the *chromatogram deconvolution* algorithm to the dataset separates each chromatogram object into its individual chromatographic peaks (Pluskal et al., 2010). In order to apply a *chromatogram deconvolution* algorithm, a *m/z* tolerance, RT tolerance, and maximum charge value need be specified.

After chromatogram deconvolution, the list of chromatographic peaks needs be subjected to an *isotope grouping* algorithm. *Isotope grouping* algorithms group peaks based on the recognition of identical isotope patterns. In order to apply an *isotope grouping* algorithm, *m/z* tolerance, RT tolerance, weight for *m/z*, and weight for RT values needs be specified. Once the peaks have been grouped based on isotope pattern, the *join aligner* algorithm is utilized to join peaks within a specified RT and *m/z* tolerance window.

If there are peaks that are only detected in a small subset of samples, they would be missing from the other samples in the dataset. These missing peaks produce missing values in the data matrix. Missing values in the data matrix greatly complexifies statistical analyses and must, therefore, be remedied (Riccadonna and Franceschi, 2018). To accomplish this task, a *gap-filling* algorithm is typically applied to the aligned data. The *gap-filling* algorithm takes the aligned data as input and re-evaluates the extracted ion chromatograms at the *m/z* and RT positions in the samples where no peak was detected (Müller et al., 2020). If a peak is detected at that point and time (where previously it was assigned a value of 0 as it had an intensity value less than the threshold specified for the *feature detection* algorithm), the value is replaced by the measured value even though it fell below the previously assigned threshold. If no value was measured, the mass feature is assigned an intensity of 0. In order to perform the *gap-filling* step (in MZMine), the *m/z* tolerance is the only parameter that needs be specified.

The final step in the pre-processing pipeline is *duplicate peak filtering*. So as to avoid the introduction of biases caused by over-representing certain prevalent mass features (i.e. to avoid

several instances of the same event being detected and skewing statistical analyses). In order to perform this step, the filter mode, *m/z* tolerance, and RT tolerance need be specified.

The end result of the data pre-processing steps should be to reduce the large amount of chromatographic and spectral raw data output by the system, down to a much smaller, statistically and computationally manageable set of peaks (Bauer et al., 2011). After subjecting the data to all of these pre-processing steps, the resulting smaller dataset may undergo (manual) data curation.

1.1.8 Statistical Analysis

Techniques of statistical analysis are utilized in metabolomics studies to extract relevant information from the data, with the aim of providing biological knowledge on the problem being examined (Saccenti et al., 2013). Just as the raw data needed to be processed according to the chosen metabolomic approach, the statistical analyses need to be informed by the chosen approach. If not, the analysis poses the risk of producing false positives and/or negatives.

1.1.9 Univariate Analysis

When only one variable is analyzed between 2 sample groups at a time, a univariate analysis is performed (Saccenti et al., 2013). Methods for univariate analysis typically include tests which identify metabolites that show statistically different patterns of detection between different sets of samples such as, *t-tests* or ANOVA (analysis of variance) (Saccenti et al., 2013). When a normal distribution and equal variance may be assumed, the Student's *t-test* may be used. In the case of most screening data, however, the sample groups will be different sizes, extracts will be of varied weight, concentration and content. As such, the sample groups have unequal variance

and a test like Welch's unequal variance *t-test* must be utilized instead. To reduce the threat (and influence) of false positives, multiple stages of testing correction must follow univariate analyses, such as the family-wise error (FEW) or false-discovery rate (FDR) method (Stelzer et al., 2013).

1.1.10 Multivariate Analysis

While as univariate analyses compare one variable at a time, multivariate analyses makes use of all variables simultaneously and considers the simultaneous relationship among the variables (Dillon and Goldstein, 1984; Saccenti et al., 2013). This is done by making correlations which reflect the extent of the relationships among the variables, in turn, identifying patterns of metabolites which differ between the examined samples groups. There are 2 general types of methods which can be utilized in multivariate analyses. They are the supervised and unsupervised methods.

In an unsupervised metabolomics experiment, all variation between metabolite patterns, regardless whether the difference in pattern may be attributed to experimental manipulations or extrinsic factors, are observed (Jansen et al., 2010). A key example of an unsupervised method which has proved useful to metabolomics may be found in the widespread use of Principal Component Analysis (PCA) (Van Der Werf et al., 2005; William Allwood et al., 2006; Jansen et al., 2010). PCA aims to capture as much variation in the dependent variables as possible and can thereby be utilized to assess clustering and outliers in the dataset based on metabolite patterns (Jansen et al., 2010; Berry et al., 2020).

Supervised methods focus on describing the subset of variation in the dependent variables which may be attributed to a predefined factor (Jansen et al., 2010). Partial Lease Squares-Discriminant Analysis (PLS-DA) is a commonly used supervised method which works to separate the variables (metabolites) into those which differ between the sample groups from those which differ for reasons unrelated to the independent variables (Jansen et al., 2010; Gromski et al., 2015). As the goal of supervised methods is to place samples in one of the defined groups, supervised methods have a tendency to be subject to overfitting the data (Zhao and Liu, 2007; Berry et al., 2020).

1.1.11 Metabolite Identification

To understand the role of each metabolite under the conditions studied, metabolomic data must be interpreted; this requires that all chemical information derived from metabolomic analyses must be linked to both biochemical causes and physiological consequences (Castle et al., 2006; Mehrotra and Mendes, 2006; Cambiaghi et al., 2016). The first step towards ultimately achieving this knowledge is in determining the identity of the metabolites. To determine the identity of the metabolites revealed in the experiment, the features may be matched to databases of known metabolites. Owing to the vastness of the metabolome and the relative infancy of metabolomics, however, a comprehensive database of metabolites produced by any non-model organism or human cell/tissue is rarely retrievable/available. As a result, in order for the metabolites revealed via an untargeted metabolomics study on *Alternaria spp*. to be identified, a database first needed be assembled. To accomplish this task, a comprehensive search of the literature was performed, with the purpose of compiling as large a list as possible, of SMs which have been detected in *Alternaria* spp. The next steps in determining the identity of the metabolite(s) would be to utilize the UV spectral data, MS/MS or fragmentation data, comparison to a reference standard and/or

isolation and characterization by NMR. These steps will be discussed in much greater detail in **Chapter 3** of the present thesis.

1.1.12 Alternaria

It is estimated that there are about 1.5 millions species that comprise the fungal kingdom (Hawksworth, 1991; O'Brien et al., 2005). Approximately 625 of those fungal species have been reported to infect vertebrates and about 8,000 species of fungi are associated with plant diseases (Fisher et al., 2021). Over the past several years, fungal species' with the capability to infect humans have caused an estimated 1.6 million deaths annually and are threatening bats, amphibians, and reptiles with extinction (Frick et al., 2010; Brown et al., 2012; Casadevall, 2017). Additionally, fungal contamination of agricultural crops causes a loss of a third of all of the crops grown annually, thereby causing billions of dollars in economic losses on an annual basis (Fisher et al., 2012; Almeida et al., 2019).

Yield losses and reduced grain quality in wheat grains are often the result of fungal contamination, before and/or after harvest (Jevtić et al., 2019). The reduction of grain quality is usually associated with reduced germination capacity and the production of harmful mycotoxins (Bălău et al., 2015). By result, the European Commission defined maximum acceptable levels for some mycotoxins in cereals and derived products intended for human/animal consumption (European Commission, 2006b, 2006a; Jevtić et al., 2019). Common contaminants of wheat grain, capable of producing mycotoxins are typically of the fungal genera *Fusarium, Aspergillus, Penicillium* and *Alternaria*.

While as many studies have focused on the management of *Fusarium* head blight and defining admissible levels of deoxynivalenol (DON) and the trichothecenes, studies dealing with the control of *Alternaria spp*. are rare (Wegulo et al., 2011; Jevtić et al., 2019). Perello and Larran, (2013) reported that colonization of wheat grain by *Alternaria* spp. at a level of 80% can reduce germination from 20 to 60% depending on the tested variety. Additionally, *Alternaria* mycotoxins have been implicated as potent mycotoxins, displaying carcinogenic, mutagenic, disruptive and/or inhibitive effects against cells and essential cellular function (Escrivá et al., 2017). It is therefore necessary to characterize and understand the different species of the genus, so when an isolate is examined, the potential risk it poses can be estimated.

Filamentous fungi are historically known as rich sources of biologically active natural products, which take the form of SMs (Soukup et al., 2016). Some filamentous fungal species have become experts at harnessing the potential to produce SMs to their advantage and, often, to the detriment of others. A prime example may be found in the genus *Alternaria*. They are known to produce a deep catalog of SMs, with functions ranging from imparting protection against predation and harsh environmental conditions, communication/signalling, competition and toxicity against other microorganisms, pathogenicity and the ability to exploit genetically-encoded weaknesses carried by various hosts (Almassi et al., 1991; Rohlfs and Churchill, 2011; Eisenman and Casadevall, 2012; Scharf et al., 2014; Fu et al., 2015; Netzker et al., 2015; Derntl et al., 2017).

The genus *Alternaria* was first established in 1816, based on the characterization of an isolate of *Alternaria tenuis* Nees, which was later identified as *Alternaria alternata* (Fr.) Keissl.

(Woudenberg et al., 2013). It has since come to encompass more than 300 different species worldwide, with lifestyles ranging from saprophytic to endophytic to biotrophic and all that's in-between (Lee et al., 2015). The majority of *Alternaria* species' lack an identified sexual stage and are therefore classified into the phylum of *Fungi Imperfecti* (otherwise called *Deuteromycetes*). Key morphological indicators of the genus *Alternaria* are in the production of large melanised conidia in chains with unique lengths and branching patterns, and with each conidium bearing a characteristic longitudinal and transverse septal pattern (Thomma, 2003). Additionally, their conidia characteristically taper at one end to an elongated beak shape (Thomma, 2003). There are, however, many exceptions to these rules and many instances where several species of different genetic background are morphologically indistinguishable from one another. As a result, the genus *Alternaria* has been fraught with taxonomic uncertainties since its inception.

Based solely on conidial size, a large portion of small-spored *Alternaria* isolates have been misidentified as *A. alternata* (Andersen et al., 2001). With other morphological features to draw on such as the three-dimensional structure of conidia and conidiophores, Simmons and Roberts (1993), delineated *Alternaria* into six different sporulation patterns, denoting six morphologically-distinct taxa. These were the first descriptions of the *A. alternata, A. tenuissima, A. cheiranthi,* and *A. brassicicola* species-groups (Simmons and Roberts, 1993; Woudenberg et al., 2015).

In an effort to achieve greater taxonomic resolution, molecular methods have been applied. While they were able to reveal that *Alternaria* species cluster in several distinct clades, they are

not able to resolve the individuals within those clades from one another (Woudenberg et al., 2013). These clades have come to be referred to as sections (Lawrence et al., 2013; Woudenberg et al., 2013, 2015). The result of these efforts has been the delineation of the genus into 30 taxonomic sections, each comprised of several species which cannot be distinguished from one another based on morphology or molecular techniques (Lawrence et al., 2013, 2015; Woudenberg et al., 2013, 2014, 2015).

1.1.13 Alternaria Secondary Metabolites

Alternaria, like many in the Dothideomycetes have the ability to produce a profound diversity of secondary metabolites (SMs). This makes them a primary food biosecurity risk as they have the ability to devastate crop yield and cause product quality issues by mycotoxin contamination (Schoch et al., 2009; Solomon, 2011). They produce SMs for various reasons, including, to penetrate plant tissue and/or cause disease symptoms in various host plant species. Several of the SMs they produce have been the subject of recent attention for the toxic properties they display. Some such examples can be found in the case of the tenuazonic acid (TeA), tentoxin (TEN), alternariol (AOH), and alternariol monomethyl ether (AME). In the realm of natural product regulation, these are often considered "emerging mycotoxins" (as they are currently being considered for regulation by the EU). The structure of these SMs may be exemplified by **Figure 1.3**, below.

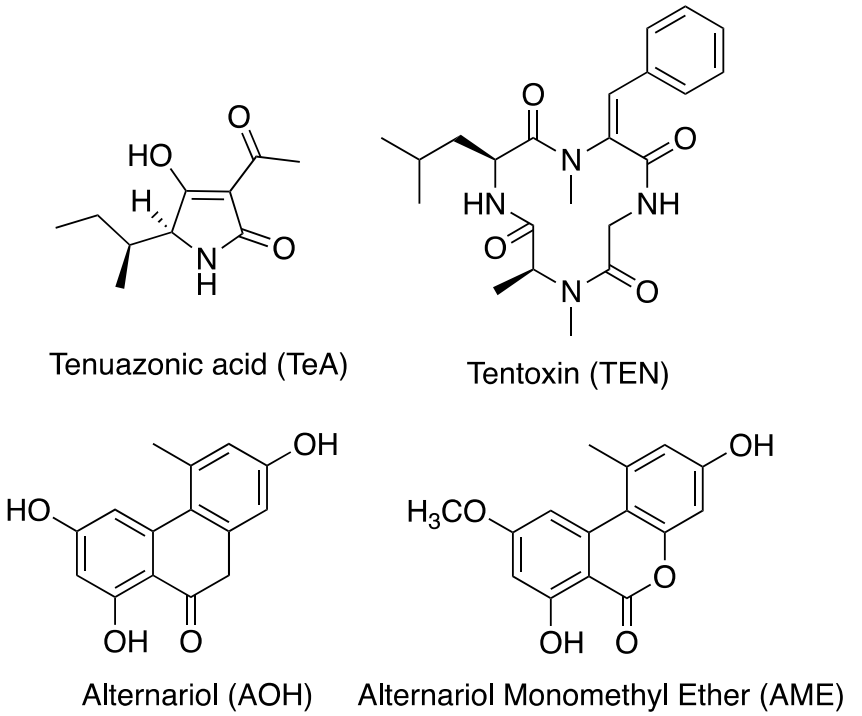

Figure 1.3. The structure of some of the secondary metabolites (SMs) produced by *Alternaria spp.*, reportedly conferring potent toxicity in human and animal cells. The structures included do not constitute a comprehensive list of all of the SMs reported in *Alternaria*, displaying potent toxicity.

If the produced SM confers an improved fitness on a wide array of host species, but is not strictly essential in order for pathogenic activity to commence/continue, it is often termed a non-host-specific-toxin (nHST). nHSTs currently encapsulates at least 30 different *Alternaria* SMs (Meena et al., 2017). Several of the more well-known examples have been evaluated by the European Food Safety Authority (EFSA) for their potential in causing risks to human health (**Figure 2.1**). For the *Alternaria* toxins displaying genotoxicity, AOH and AME, a high level of concern was expressed, outlining a need for additionally toxicity data (EFSA Panel on Contaminants in the Food Chain (CONTAM), 2011). For the non-genotoxic TEN and TeA, it was initially decided that they are unlikely to be a concern to human health (EFSA Panel on Contaminants in the Food Chain (CONTAM), 2011).

While as studies since conducted in Europe, quantifying the presence of TeA in random food samples from grocery stores, has increased the pressure on EFSA to impose limits/screen against TeA, information regarding its toxicology remains very limited (Kumari and Tirkey, 2019). Such is also the case for many other nHSTs reported in *Alternaria spp*., such as, zinniol, brefeldin A, curvularin, dehydrocurvularin, alterotoxin (ATX) I, II and III, which have all been characterized and are considered potent mycotoxins, though the risk they pose to human health, their modes of action and role in virulence largely remains unknown (Andersen et al., 2015; Lee et al., 2015).

Some of the SMs produced by *Alternaria* have a function which is absolutely required in order for some pathogenic *Alternaria* to invade host tissue and cause disease symptoms in susceptible hosts. SMs displaying these characteristics have been found to display exclusive toxicity to particular host species', therefore, earning them the term of Host-specific toxins (HSTs). To date, 20 different HSTs which have been identified in *Alternaria*. The susceptibility of the host to the HST(s) is genotype-derived and studies have come to show that in the majority of cases, sensitivity to the toxin is a dominant trait, implying that the targeted host gene product is the direct or indirect receptor of the toxin (Friesen et al., 2008). In *Alternaria*, the genes necessary for the biosynthesis of HSTs have consistently and exclusively been localized as clusters of tightly regulated genes on areas of "extra genome or chromosomal content".

1.1.14 Accessory Chromosomes (ACs)

ACs are chromosomes that are carried by some pathogenic *Alternaria* and are comprised of "extra" genetic content (i.e. genetic information which is unique from the core genome). ACs are generally small (less than 2.0Mb), dispensable for growth and believed to be transmitted between

isolates of a population through horizontal gene transfer (HGT). They are considered dispensable for growth as have only been found in pathogenic isolates and, when lost, those isolates only lose their pathogenicity towards the certain host species but can otherwise function like normal. With the advent of long-read genome sequencing, the ability to identify ACs from core chromosomes has improved dramatically in recent years.

The origin of ACs in *Alternaria* remains unknown so Tsuge et al., (2016) went about comparing the entire lengths of ACs from the strawberry, tomato, and apple pathotypes of *A. alternata*. In doing so, they identified large syntenic regions among the three, but completely unique HST-biosynthesis gene clusters. They therefore proposed that the syntenic regions are the core of the original extra chromosome, and the clusters responsible for HST-biosynthesis were integrated into the original chromosome by HGT. Their origin remains a controversial topic.

With the thought in mind that ACs are rare amongst a sample of given *Alternaria* isolates, an untargeted metabolomics approach offers a unique opportunity for screening a population of isolates for the production of unique secondary metabolites within the population that could be attributed to the presence of ACs or accessory regions of core chromosomes.

1.2 Objectives

The overall objective of the present thesis was to use untargeted LC-MS metabolomics to obtain a robust profile for 60 strains of *Alternaria spp.*, selected from various species-groups in the *Alternaria* genus. In turn, a massive dataset would be produced which would be utilized to reveal trends that further our understanding of the natural products they produce.

The specific hypotheses of the present thesis were as follows:

Chapter 2: That mass features displaying a unique and isolated production pattern indicated the activation of biosynthetic gene clusters (BGCs) that could be localized to ACs (or have other interesting implications).

Chapter 3: That mass features showing unique production patterns in the *alternata* group studied in **Chapter 2**, are a family of compounds/analogs, sharing structural similarities.

Chapter 4: That the family of compounds/analogs obtained in **Chapter 3** are produced by the dehydrocurvularin (DHC) BGC, that is shared by DET2035, KAS5321, and KAS5743. Additionally, that this BGC can be implicated as being shared amongst different fungal plant pathogens.

Towards proving or disproving these hypotheses, the specific aims of the experiments were:

Chapter 2: Acquire a robust SM profile for different *Alternaria spp*. for a selection of *A. sect. A. alternata* and *A. sect. infectoria* isolates, to assess for the potential presence of SMs associated with ACs (or having other interesting implications).

Chapter 3: Perform a scale-up growth fermentation of isolates having unique mass features with the intent to isolate, purify, and elucidate the identity of the metabolites corresponding to those mass features.

Chapter 4: Identify and characterize the BGCs in the genomes of DET2035, KAS5321, and KAS5743 that correspond with the structural motifs of the metabolites identified in Chapters 3 and to assess the relatedness of this BGC with other fungi.

1.3 References

Almassi, F., Ghisalberti, E. L., Narbey, M. J., and Sivasithamparam, K. (1991). New antibiotics from strains of trichoderma harzianum. *J. Nat. Prod.* 54, 396–402. doi:10.1021/np50074a008.

Almeida, F., Rodrigues, M. L., and Coelho, C. (2019). The Still Underestimated Problem of Fungal Diseases Worldwide . *Front. Microbiol.* 10, 214. Available at: https://www.frontiersin.org/article/10.3389/fmicb.2019.00214.

Andersen, B., KrØger, E., and Roberts, R. G. (2001). Chemical and morphological segregation of *Alternaria* alternata, A. Gaisen and A. Longipes. *Mycol. Res.* 105, 291–299. doi:10.1017/S0953756201003446.

Andersen, B., Nielsen, K. F., Fernández Pinto, V., and Patriarca, A. (2015). Characterization of *Alternaria* strains from Argentinean blueberry, tomato, walnut and wheat. *Int. J. Food Microbiol.* doi:10.1016/j.ijfoodmicro.2014.11.029.

Bălău, A.-M., Lipșa, F.-D., and Ulea, E. (2015). The Fungal Load Evaluation Of Some Wheat Varieties Kernels In Storage Conditions. *Lucr. Ştiinţifice* 58, 93–96.

Bartee, L., Shriner, W., and Creech, C. (2017). *Principles of Biology*. Open Oregon Educational Resources Available at: https://openoregon.pressbooks.pub/mhccmajorsbio/ [Accessed August 22, 2021].

Barton, R. H., Nicholson, J. K., Elliott, P., and Holmes, E. (2008). High-throughput 1H NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: Validation study. *Int. J. Epidemiol.* 37. doi:10.1093/ije/dym284.

Bauer, C., Cramer, R., and Schuchhardt, J. (2011). Evaluation of peak-picking algorithms for protein mass spectrometry. *Methods Mol. Biol.* 696, 341–352. doi:10.1007/978-1-60761-

987-1_22.

Berry, M. W., Mohamed, A., and Yap, B. W. (2020). *Supervised and Unsupervised Learning for Data Science*. Cham: Springer International Publishing doi:10.1007/978-3-030-22475-2.

Brown, G. D., Denning, D. W., Gow, N. A. R., Levitz, S. M., Netea, M. G., and White, T. C. (2012). Hidden Killers: Human Fungal Infections. *Sci. Transl. Med.* 4, 165rv13-165rv13. doi:10.1126/SCITRANSLMED.3004404.

Büscher, J. M., Czernik, D., Ewald, J. C., Sauer, U., and Zamboni, N. (2009). Cross-platform comparison of methods for quantitative metabolomics of primary metabolism. *Anal. Chem.* 81, 2135–2143. doi:10.1021/ac8022857.

Cambiaghi, A., Ferrario, M., and Masseroli, M. (2016). Analysis of metabolomic data: tools, current strategies and future challenges for omics data integration. *Brief. Bioinform.* 18, 498–510. doi:10.1093/bib/bbw031.

Casadevall, A. (2017). Don't Forget the Fungi When Considering Global Catastrophic Biorisks. *https://home.liebertpub.com/hs* 15, 341–342. doi:10.1089/HS.2017.0048.

Castle, A. L., Fiehn, O., Kaddurah-Daouk, R., and Lindon, J. C. (2006). Metabolomics Standards Workshop and the development of international standards for reporting metabolomics experimental results. *Brief. Bioinform.* 7, 159–165. doi:10.1093/bib/bbl008.

Cooper, G. M. (2000). Chromosomes and Chromatin. Available at: https://www.ncbi.nlm.nih.gov/books/NBK9863/ [Accessed August 22, 2021].

Derntl, C., Kluger, B., Bueschl, C., Schuhmacher, R., Mach, R. L., and Mach-Aigner, A. R. (2017). Transcription factor Xpp1 is a switch between primary and secondary fungal metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 114, E560–E569. doi:10.1073/pnas.1609348114.

Dillon, W. R., and Goldstein, M. (1984). *Multivariate Analysis: Methods and Applications | Wiley*. Available at: https://www.wiley.com/en-us/Multivariate+Analysis%3A+Methods+and+Applications-p-9780471083177 [Accessed January 8, 2021].

Du, P., Kibbe, W. A., and Lin, S. M. (2006). Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics* 22, 2059–2065. doi:10.1093/bioinformatics/btl355.

Dudley, E., Yousef, M., Wang, Y., and Griffiths, W. J. (2010). "Chapter 2 - Targeted metabolomics and mass spectrometry," in, ed. R. B. T.-A. in P. C. and S. B. Donev (Academic Press), 45–83. doi:https://doi.org/10.1016/B978-0-12-381264-3.00002-3.

EFSA Panel on Contaminants in the Food Chain (CONTAM), E. (2011). Scientific opinion on the risks for animal and public health related to the presence of *Alternaria* toxins in feed and food. *EFSA J.* 9, 2407. doi:https://doi.org/10.2903/j.efsa.2011.2407.

Eisenman, H. C., and Casadevall, A. (2012). Synthesis and assembly of fungal melanin. *Appl. Mirobiol. Biotechnol.* 93, 931–940. doi:10.1007/s00253-011-3777-2.

Erb, M., and Kliebenstein, D. J. (2020). Plant Secondary Metabolites as Defenses, Regulators, and Primary Metabolites: The Blurred Functional Trichotomy. *Plant Physiol.* 184, 39–52. doi:10.1104/pp.20.00433.

Escrivá, L., Oueslati, S., Font, G., and Manyes, L. (2017). *Alternaria* Mycotoxins in Food and Feed: An Overview. *J. Food Qual.* 2017, 1569748. doi:10.1155/2017/1569748.

European Commission (2006a). Commission recommendation on the presence of deoxynivalenol, zearalenone, ochratoxin A, T-2 and HT-2 and fumonisins in products intended for animal feeding (Text with EEA relevance) At the request of the Commission

the European Food Safety Authority (EFSA. *Off. J. Eur. Union* 229, 7–9. Available at: http://www.efsa.europa.eu/etc/medialib/efsa/science/contam/ [Accessed July 18, 2021].

European Commission (2006b). Commission regulation (EC) no. 1881/2006 of 19 December 2006 Setting maximum levels of certain contaminants in foodstuffs. *Off. J. Eur. Union* 364, 5–24.

Fisher, M. C., Gurr, S. J., Cuomo, C. A., Blehert, D. S., Jin, H., Stukenbrock, E. H., et al. (2021). Threats Posed by the Fungal Kingdom to Humans, Wildlife, and Agriculture. *MBio* 11, e00449-20. doi:10.1128/mBio.00449-20.

Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., et al. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484, 186–194. doi:10.1038/nature10947.

Frick, W. F., Pollock, J. F., Hicks, A. C., Langwig, K. E., Reynolds, D. S., Turner, G. G., et al. (2010). An Emerging Disease Causes Regional Population Collapse of a Common North American Bat Species. *Science (80-.).* 329, 679 LP – 682. doi:10.1126/science.1188594.

Friesen, T. L., Faris, J. D., Solomon, P. S., and Oliver, R. P. (2008). Host-specific toxins: effectors of necrotrophic pathogenicity. *Cell. Microbiol.* 10, 1421–1428. doi:10.1111/j.1462-5822.2008.01153.x.

Fu, S.-F., Wei, J.-Y., Chen, H.-W., Liu, Y.-Y., Lu, H.-Y., and Chou, J.-Y. (2015). Indole-3-acetic acid: A widespread physiological code in interactions of fungi with other organisms. *Plant Signal. Behav.* 10, e1048052–e1048052. doi:10.1080/15592324.2015.1048052.

Gardner, M. W., and Brodbelt, J. S. (2009). Reduction of Chemical Noise in Electrospray Ionization Mass Spectrometry by Supplemental IR Activation. *J. Am. Soc. Mass Spectrom.* 20, 2206–2210. doi:https://doi.org/10.1016/j.jasms.2009.08.008.

Genome (2014). *Nat. Educ.* Available at: https://www.nature.com/scitable/definition/genome-43/ [Accessed August 22, 2021].

Griffin, J. L., Nicholls, A. W., Keun, H. C., Mortishire-Smith, R. J., Nicholson, J. K., and Kuehn, T. (2002). Metabolic profiling of rodent biological fluids via 1H NMR spectroscopy using a 1 mm microlitre probe. *Analyst* 127, 582–584. doi:10.1039/b201207c.

Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., et al. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding. *Anal. Chim. Acta* 879, 10–23. doi:10.1016/j.aca.2015.02.012.

Hawksworth, D. L. (1991). The fungal dimension of biodiversity: magnitude, significance, and conservation. *Mycol. Res.* 95, 641–655. doi:10.1016/S0953-7562(09)80810-1.

Irizarry, R. A. (2019). *Introduction to data science : data analysis and prediction algorithms with R*. Chapman and Hall/CRC.

Jansen, J. J., Smit, S., Hoefsloot, H. C. J., and Smilde, A. K. (2010). The photographer and the greenhouse: How to analyse plant metabolomics data. *Phytochem. Anal.* 21, 48–60. doi:10.1002/pca.1181.

Jevtić, R., Župunski, V., Lalošević, M., and Živanov, S. T. (2019). Colonization of winter wheat grain with *Fusarium and Alternaria* species and influence on pest control management. *J. Gen. Plant Pathol.* 85, 273–281. doi:10.1007/s10327-019-00844-y.

Johnson, C. H., Ivanisevic, J., and Siuzdak, G. (2016). Metabolomics: Beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* 17, 451–459. doi:10.1038/nrm.2016.25.

Jonsson, P., Bruce, S. J., Moritz, T., Trygg, J., Sjöström, M., Plumb, R., et al. (2005). Extraction, interpretation and validation of information for comparing samples in metabolic LC/MS

data sets. *Analyst* 130, 701–707. doi:10.1039/b501890k.

Kell, D. B., and Oliver, S. G. (2016). The metabolome 18 years on: a concept comes of age. *Metabolomics* 12, 148. doi:10.1007/s11306-016-1108-4.

Kumari, A., and Tirkey, N. N. (2019). Tenuazonic Acid: A Potent Mycotoxin. *Recent Trends Hum. Anim. Mycol.*, 203–211. doi:10.1007/978-981-13-9435-5_8.

Kwon, D., Vannucci, M., Joon, J. S., Jeong, J., and Pfeiffer, R. M. (2008). A novel wavelet-based thresholding method for the pre-processing of mass spectrometry data that accounts for heterogeneous noise. *Proteomics* 8, 3019–3029. doi:10.1002/pmic.200701010.

Lange, E., Gropl, C., Reinert, K., Kohlbacher, O., and Hildebrandt, A. (2006). High-accuracy peak picking of proteomics data using wavelet techniques. in *Proceedings of the Pacific Symposium on Biocomputing 2006, PSB 2006* (Institute of Computer Science, Free University of Berlin Takustr. 9, 14195 Berlin, Germany. lange@inf.fu-berlin.de FAU - Gropl, Clemens), 243–254.

Lawrence, D. P., Gannibal, P. B., Peever, T. L., and Pryor, B. M. (2013). The sections of *Alternaria*: formalizing species-group concepts. *Mycologia* 105, 530–546. doi:10.3852/12-249.

Lawrence, D. P., Rotondo, F., and Gannibal, P. B. (2015). Biodiversity and taxonomy of the pleomorphic genus *Alternaria*. *Mycol. Prog.* 15, 3. doi:10.1007/s11557-015-1144-x.

Lee, H. B., Patriarca, A., and Magan, N. (2015). *Alternaria* in Food: Ecophysiology, Mycotoxin Production and Toxicology. *Mycobiology* 43, 93–106. doi:10.5941/MYCO.2015.43.2.93.

Lengyel, P. (1969). The Process of Translation as Seen in 1969. *Cold Spring Harb. Symp. Quant. Biol.* 34, 828–841. doi:10.1101/SQB.1969.034.01.094.

Liu, M. A. (2019). A Comparison of Plasmid DNA and mRNA as Vaccine Technologies.

Vaccines 2019, Vol. 7, Page 37 7, 37. doi:10.3390/VACCINES7020037.

Lodish, H., Berk, A., and Zipursky, S. (2000). "Cell Biology. 4th edition," in *Chapter 1: The Dynamic Cell. of Molecular Cell Biology*.

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., and Shafee, T. (2017). Transcriptomics technologies. *PLoS Comput. Biol.* 13, e1005457–e1005457. doi:10.1371/JOURNAL.PCBI.1005457.

Meena, M., Gupta, S. K., Swapnil, P., Zehra, A., Dubey, M. K., and Upadhyay, R. S. (2017). *Alternaria* toxins: Potential virulence factors and genes related to pathogenesis. *Front. Microbiol.* 8, 1451. doi:10.3389/fmicb.2017.01451.

Mehrotra, B., and Mendes, P. (2006). "Bioinformatics approaches to integrate metabolomics and other systems biology data," in *Biotechnology in Agriculture and Forestry* Biotechnology in Agriculture and Forestry., eds. R. A. Dixon and L. Willmitzer (Berlin, Heidelberg: Springer Berlin Heidelberg), 105–115. doi:10.1007/3-540-29782-0_8.

Müller, E., Huber, C., Beckers, L.-M., Brack, W., Krauss, M., and Schulze, T. (2020). A Data Set of 255,000 Randomly Selected and Manually Classified Extracted Ion Chromatograms for Evaluation of Peak Detection Methods. *Metabolites* 10, 162. doi:10.3390/metabo10040162.

Netzker, T., Fischer, J., Weber, J., Mattern, D. J., König, C. C., Valiante, V., et al. (2015). Microbial communication leading to the activation of silent fungal secondary metabolite gene clusters. *Front. Microbiol.* 6, 299. doi:10.3389/fmicb.2015.00299.

Nordström, A., Want, E., Northen, T., Lehtiö, J., and Siuzdak, G. (2008). Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics. *Anal. Chem.* 80, 421–429. doi:10.1021/ac701982e.

O'Brien, H. E., Parrent, J. L., Jackson, J. A., Moncalvo, J.-M., and Vilgalys, R. (2005). Fungal community analysis by large-scale sequencing of environmental samples. *Appl. Environ. Microbiol.* 71, 5544–5550. doi:10.1128/AEM.71.9.5544-5550.2005.

Patti, G. J., Yanes, O., and Siuzdak, G. (2012). Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* 13, 263–269. doi:10.1038/nrm3314.

Perello, A. E., and Larran, S. (2013). Nature and effect of *Alternaria* spp. complex from wheat grain on germination and disease transmission. *Pakistan J. Bot.* 45, 1817–1824.

Pluskal, T., Castillo, S., Villar-Briones, A., and Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11, 395. doi:10.1186/1471-2105-11-395.

Ribbenstedt, A., Ziarrusta, H., and Benskin, J. P. (2018). Development, characterization and comparisons of targeted and non-targeted metabolomics methods. *PLoS One* 13, e0207082. Available at: https://doi.org/10.1371/journal.pone.0207082.

Riccadonna, S., and Franceschi, P. (2018). Data treatment for LC-MS untargeted analysis. *Methods Mol. Biol.* 1738, 27–39. doi:10.1007/978-1-4939-7643-0_3.

Rinschen, M. M., Ivanisevic, J., Giera, M., and Siuzdak, G. (2019). Identification of bioactive metabolites using activity metabolomics. *Nat. Rev. Mol. Cell Biol.* 20, 353–367. doi:10.1038/s41580-019-0108-4.

Roberts, L. D., Souza, A. L., Gerszten, R. E., and Clish, C. B. (2012). Targeted metabolomics. *Curr. Protoc. Mol. Biol.* 1. doi:10.1002/0471142727.mb3002s98.

Rohlfs, M., and Churchill, A. C. L. (2011). Fungal secondary metabolites as modulators of interactions with insects and other arthropods. *Fungal Genet. Biol.* 48, 23–34. doi:10.1016/j.fgb.2010.08.008.

Saccenti, E., Hoefsloot, H., Smilde, A., Westerhuis, J., and Hendriks, M. (2013). Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* 10, e1-15. doi:10.1007/s11306-013-0598-6.

Scharf, D. H., Heinekamp, T., and Brakhage, A. A. (2014). Human and Plant Fungal Pathogens: The Role of Secondary Metabolites. *PLoS Pathog*. 10, e1003859. doi:10.1371/journal.ppat.1003859.

Schoch, C. L., Crous, P. W., Groenewald, J. Z., Boehm, E. W. A., Burgess, T. I., de Gruyter, J., et al. (2009). A class-wide phylogenetic assessment of Dothideomycetes. *Stud. Mycol.* 64, 1–15. doi:10.3114/sim.2009.64.01.

Simmons, E. G., and Roberts, R. G. (1993). *Alternaria* themes and variations (73). *Mycotaxon* 48, 109–140. Available at: https://www.cabi.org/isc/abstract/19942300558 [Accessed January 21, 2021].

Solomon, P. S. (2011). Assessing the mycotoxigenic threat of necrotrophic pathogens of wheat. *Mycotoxin Res.* 27, 231–237. doi:10.1007/s12550-011-0108-5.

Soukup, A. A., Keller, N. P., and Wiemann, P. (2016). "Enhancing nonribosomal peptide biosynthesis in filamentous fungi," in *Methods in Molecular Biology*, ed. B. S. Evans (New York, NY: Springer New York), 149–160. doi:10.1007/978-1-4939-3375-4_10.

Stelzer, J., Chen, Y., and Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *Neuroimage* 65, 69–82. doi:10.1016/j.neuroimage.2012.09.063.

Thomma, B. P. H. J. (2003). *Alternaria* spp.: From general saprophyte to specific parasite. *Mol. Plant Pathol.* 4, 225–236. doi:10.1046/j.1364-3703.2003.00173.x.

Tsuge, T., Harimoto, Y., Hanada, K., Akagi, Y., Kodama, M., Akimitsu, K., et al. (2016).

Evolution of pathogenicity controlled by small, dispensable chromosomes in *Alternaria* alternata pathogens. *Physiol. Mol. Plant Pathol.* 95, 27–31. doi:10.1016/J.PMPP.2016.02.009.

Van Der Werf, M. J., Jellema, R. H., and Hankemeier, T. (2005). Microbial metabolomics: Replacing trial-and-error by the unbiased selection and ranking of targets. *J. Ind. Microbiol. Biotechnol.* 32, 234–252. doi:10.1007/s10295-005-0231-4.

Vining, L. C. (1992). Secondary metabolism, inventive evolution and biochemical diversity-a review* (Intrinsic and extrinsic functions; phenotypic.

Wegulo, S. N., Bockus, W. W., Nopsa, J. H., de Wolf, E. D., Eskridge, K. M., Peiris, K. H. S., et al. (2011). Effects of integrating cultivar resistance and fungicide application on *Fusarium* head blight and deoxynivalenol in winter wheat. *Plant Dis.* 95, 554–560. doi:10.1094/PDIS-07-10-0495.

William Allwood, J., Ellis, D. I., Heald, J. K., Goodacre, R., and Mur, L. A. J. (2006). Metabolomic approaches reveal that phosphatidic and phosphatidyl glycerol phospholipids are major discriminatory non-polar metabolites in responses by Brachypodium distachyon to challenge by Magnaporthe grisea. *Plant J.* 46, 351–368. doi:10.1111/j.1365-313X.2006.02692.x.

Woudenberg, J. H. C., Groenewald, J. Z., Binder, M., and Crous, P. W. (2013). *Alternaria* redefined. *Stud. Mycol.* 75, 171–212. doi:10.3114/sim0015.

Woudenberg, J. H. C., Seidl, M. F., Groenewald, J. Z., de Vries, M., Stielow, J. B., Thomma, B. P. H. J., et al. (2015). *Alternaria* section *Alternaria*: Species, *formae speciales* or pathotypes? *Stud. Mycol.* 82, 1–21. doi:10.1016/j.simyco.2015.07.001.

Woudenberg, J. H. C., Truter, M., Groenewald, J. Z., and Crous, P. W. (2014). Large-spored

Alternaria pathogens in section Porri disentangled. *Stud. Mycol.* 79, 1–47. doi:10.1016/j.simyco.2014.07.003.

Zamboni, N., Saghatelian, A., and Patti, G. J. (2015). Defining the metabolome: size, flux, and regulation. *Mol. Cell* 58, 699–706. doi:10.1016/j.molcel.2015.04.021.

Zhang, H.-W., Lv, C., Zhang, L.-J., Guo, X., Shen, Y.-W., Nagle, D. G., et al. (2021). Application of omics- and multi-omics-based techniques for natural product target discovery. *Biomed. Pharmacother.* 141, 111833. doi:https://doi.org/10.1016/j.biopha.2021.111833.

Zhao, Z., and Liu, H. (2007). Spectral feature selection for supervised and unsupervised learning. *ACM Int. Conf. Proceeding Ser.* 227, 1151–1157. doi:10.1145/1273496.1273641.

Chapter 2

2 Chemical Profiles as a Tool for Unique Secondary Metabolite Discovery

2.1 Introduction

As discussed above, the genus *Alternaria* has been fraught with taxonomic uncertainty since its inception. A precise and correct identification of the species is necessary, not only because of the human desire to classify and control, but because the species name embodies a set of characteristics which allow for the prediction of its behaviour (Andersen et al., 2001). Based upon morphology, Simmons and Roberts, (1993) were able to delineate the genera into 6 morphological species-groups. In an effort to achieve greater resolution within the genera, molecular methods have been employed. The result of these efforts was the delineation of the genera into 30 taxonomic sections, each comprised of several species (Lawrence et al., 2013, 2015; Woudenberg et al., 2013, 2014, 2015). Due to a lack of molecular variation, a molecular study pooled the *Alternaria arborescens* and *Alternaria tenuissima* species-groups with *Alternaria alternata* into one section, now comprised of more than 50 species (Andrew et al., 2009; Lawrence et al., 2013; Woudenberg et al., 2013; Andersen et al., 2015).

Chemotaxonomy is "the classification and identification of filamentous fungi based on profiles of secondary metabolites" (Frisvad et al., 2008). The production of SMs, including both known and unknown compounds, on standardized laboratory media has been successfully utilized in metabolomics-based chemotaxonomic studies in large ascomycete genera like *Aspergillus*,

Fusarium, Hypoxylon, Penicillium, Stachybotrys, Xylaria and in few basidiomycete genera (Andersen et al., 2001; Frisvad et al., 2008). A chemotaxonomic approach to classifying *Alternaria spp*. has been attempted on numerous occasions. While as a chemotaxonomic approach has proven effective at resolving *Alternaria infectoria* from other small-spored and/or large-spored *Alternaria spp.*, it has been proven to be largely ineffective at resolving relationships within and between other *Alternaria spp*. (Andersen and Thrane, 1996; Andersen et al., 2002, 2008, 2009).

A study conducted by Andersen et al. (2015), for instance, attempted to utilize the morphology and/or metabolite profiles to classify 87 Argentinian strains of *Alternaria*. Utilization of the morphology and/or metabolite profiles proved to be unable to assign any particular species; only to species-groups (Andersen et al., 2015). Additionally, a study conducted by Kelman et al. (2020) examined the chemical profile of 128 strains of Canadian strains of *Alternaria*. In doing so, they were able to distinguish *A. infectoria* from the other strains examined but were otherwise unable to resolve *Alternaria* at the species-level (Kelman et al., 2020). In both reports, the inability to utilize metabolite profiles to classify *Alternaria spp*. could largely be attributed to inconsistent patterns of metabolite production within the different species-groups. A factor that may be contributing to the observed differences in SM production patterns, could be the presence of unique SM BGCs occurring on ACs and in accessory regions (ARs) on core chromosomes in many *Alternaria* genomes (Andersen et al., 2015; da Cruz Cabral et al., 2017; Patriarca et al., 2019; Somma et al., 2019; Kelman et al., 2020).

Within the species *A. alternata,* the terms *forma specialis (f.sp.)* and pathotype have been used (interchangeably) to describe isolates which are morphologically indistinguishable but infect specific hosts. Nishimura and Kohmoto (1983) proposed that *Alternaria* strains which are morphologically identical but produce different HSTs (and thereby have distinct host ranges), should be defined as different pathotypes of *A. alternata*. Currently, this would include seven different pathotypes of *A. alternata,* and 16 different *f. sp*. epithets (Akimitsu et al., 2014; Woudenberg et al., 2015).

Gene clusters responsible for the biosynthesis of HSTs in the various *A. alternata f. sp.*/pathotypes reported to date have only been found to occur on ACs (Witte et al., 2021). The genes which encode many of these toxins have been studied at-length, however, the genomic content and organization of ACs remains largely unknown (Hu et al., 2012). Recently, a study by Wang et al. (2019), isolated and sequenced the ACs of a strain of *A. alternata* (Z7). The ACs were predicted to contain 254 protein-coding genes, predominantly having functional categories associated with "metabolic processes" and having relatively few genes associated with CAZymes, transporters and kinases (Wang et al., 2019). In a study conducted by Hu et al., (2012) the AC of an *A. arborescens* isolate was sequenced. In this study, they screened each AC gene and those surrounding them, looking for evidence of clustering of PKS, NRPS, transcription factors, transporters, P450 proteins, FAD binding proteins, transferases, and oxidoreductases. By result, they were able to identify 29 PKS, 5 NRPS, and 2 hybrid PKS-NRPS genes on the AC, comprising 10 putative SM gene clusters. As studies with the aim of discovering and characterizing the other SM gene clusters on ACs are rare, there is a need to fill that gap in our understanding of ACs in *Alternaria*.

The synthesis of SMs is tightly controlled to avoid production and, therefore, resource consumption under conditions where they are not needed, hence why large-scale sequencing data has been able to show that the clustering of genes necessary for SM biosynthesis is the rule and not the exception (Gacek and Strauss, 2012). As the genes are often physically linked on the chromosome, they are also often transcriptionally co-regulated. By achieving the conditions necessary to promote the activation of a given SM gene, the entire gene cluster often accompanies. Such conditions could include when macro- or micro-nutrients become limiting or when stress is imposed on the cells like by humidity, temperature, UV irradiation, salt, or unfavourable pH values challenge the physiological integrity of the cells (Yu and Keller, 2005; Hoffmeister and NP, 2007; Gacek and Strauss, 2012). The growth conditions used in this experiment were selected to specifically stimulate SM production through inducing a stress response. The light and temperature conditions were meant to mimic the natural conditions which *Alternaria* would be subject to in North America and Europe.

CYS80 contains corn meal, yeast extract and sucrose (80g per 1L) in solution. This means that CYS80 has a high starch and sugar content. Sucrose is stressful to fungal organisms as it is the predominant nutrient source in the media, though it has been shown that most fungi have difficulties metabolizing sucrose, if they have the ability to use it at all (Greene et al., 1953). Corn meal is traditionally utilized as it promotes the development of various microscopic morphological structures that aid in the identification of the sample grown upon it (Hazen and Reed, 1955). DRYES contains dichloran (2,6-dichloro-4-nitroaniline), rose bengal, yeast extract, and sucrose. DRYES is therefore a stressful media as several of the ingredients are considered potent toxins. Dichloran, alone and in combination with rose Bengal, has been shown to inhibit

the growth of several fungal genera, while limiting the colony diameters of other genera (King et al., 1979; Hocking and Pitt, 1980). Rose bengal has been shown to be cytotoxic to bacteria, and to inhibit the growth of many fungal species, while selecting for others (Tsao, 1970; King et al., 1979; Banks et al., 1985). In this way, the use of DRYES both ensures that contamination is mitigated, while inducing the activation of BGCs that would typically remain silent, if not for activating the fungal stress response.

MMK2 primarily contains mannitol (as the sole carbon source), Murashige and Skoog salts, and KH_2PO_4, and $MgSO_4 \cdot 7H_2O$. The use of these ingredients induces a stress response in fungi for several reasons. Mannitol production and secretion is essential for the pathogenicity of several fungal pathogens of both animals and plants (Chaturvedi et al., 1996a, 1996b; Meena et al., 2015). A study by Jennings et al., (1998) revealed that when *A. alternata* infects tobacco, it secretes mannitol to quench the ROS that mediate plant defenses. By including mannitol in the media, the fungi are immediately provided with the signal to assume a pathogenic lifestyle, in turn, activating as many BGCs as possible. The use of Murashige and Skoog salts in the media was also to serve the purpose of inducing a stress response in the fungi. Specifically, its use was to induce salt stress, while also mimicking the problem of soil salinity, commonly faced by crop farmers (Shoaib et al., 2018).

ZM2 was selected as a fermentation media for its relative lack of nutrients. In that the only sugar sources in the media are complex carbohydrates derived from whole oats and a small concentration of molasses, fungi are forced to express various extracellular enzyme to breakdown these complex plant-based polymers. These enzymes are the same as are used to

degrade plant cell walls and sequester nutrients during infection. As such, by promoting their expression, genes conferring pathogenic activity might be activated as well.

Historically, the application of metabolomics to the *Alternaria* genus has been with the intent to achieve taxonomic resolution. As previously described, there are often important differences in the metabolomes of the isolates within the same species, confounding efforts to classify populations within a species based on "core"-associated genotypes (Witte et al., 2021). Combining genomics and untargeted metabolomics in a "multi-omics" strategy, however, has provided an extremely powerful way to screen populations of fungi for SM production associated with interesting events, such as Accessory Chromosome (AC) detection, or detection of poorly understood virulence factors (Witte et al., 2021). The application of genomics allows a "top-down" approach where high-quality genome sequences are used to characterize and contextualize fungal genomes at the chromosomal level (Witte et al., 2021). Untargeted metabolomic analysis, by contrast, affords a "bottom-up" approach, where chemical extracts are analyzed using high-resolution mass spectrometry to characterize and compare patterns in SM production observed within a population to search for unique patterns attributable to sub-sets within a population (Witte et al., 2021). Leveraging results obtained from both genomic and metabolomic experimentation within a species population allows for correlation of specific SMs to specific BGCs residing in the fungal genome (or AC/AR) (Witte et al., 2021).

Continuing advancements in the fields of high-resolution mass spectrometry (HRMS), metabolite databases, and data analysis have propelled the application of metabolomics to the forefront of SM screening in filamentous fungi (Kelman et al., 2020). Untargeted metabolomic

approaches can be utilized to characterize and compare *en masse* fungal culture extracts for both known and unknown fungal metabolites, using visual representations of SM expression generated from untargeted liquid chromatography (HPLC) – HRMS analyses (Andersen et al., 2001; Kelman et al., 2020). Visual representations of SM expression have been utilized previously to identify the presence of metabolites which could be linked to ACs (Witte et al., 2020). In this study, the biosynthetic potential of 38 *Fusarium poae* isolates from Eastern Canada was examined using a combination of long-read and short-read genome sequencing and untargeted, HRMS metabolomic analysis on the extracts from isolates cultured in multiple media conditions (Witte et al., 2020). As a result, they were able to discover the BGC of the apicidins on an AC (Witte et al., 2020). Additionally, untargeted metabolomics techniques have recently proven to be capable of screening for biomarkers of aflatoxigenic *Aspergillus* species from field samples taken in China (Xie et al., 2022).

The experiment of the present chapter was to acquire the SM profile of 60 uncharacterized strains of *Alternaria* (representing a selection of *A. alternata* and *A. infectoria* isolates) and to then compare the SM expression profiles of those 60 isolates to assess for the presence of SMs with interesting effects/implications. Four culture media were selected to promote the maximal production of SMs from each of the selected isolates. Each culture was extracted and profiled for SM production. From the accompanying untargeted metabolomics data analysis of the in vitro SM production, unique mass features were selected for future molecular identification and characterization in **Chapter 3** of the present thesis.

2.2 Materials and Methods

2.2.1 Media, Chemicals, and Solvents

The 4 media conditions utilized for the work of the present chapter, were CYS80, DRYES, MMK2, and ZM2. The specifics of their formulations may be found in the **Media Formulations** section of **Appendix A** of the present chapter. The chemicals and solvents utilized for the experiments conducted in the present chapter, are detailed in **Table 2.2** of the **Chemicals and Solvents** section of **Appendix A**.

2.2.2 Fungal Strains

A total of 59 fungal isolates previously identified as *Alternaria* were utilized for the present study, along with 1 *Pithomyces* isolate. Previous identification work was performed by a collaborating lab group (Dr. Jeremy Dettman of the Agriculture and Agri-Food Canada, Ottawa Research and Development Centre), whereby, DNA barcoding genes previously found to be reliable for *Alternaria* molecular taxonomy were utilized to classify which species-group the given isolate(s) belong to. A comprehensive list of the strains selected and the section/species-group to which they belong may be found in **Table 6.1** of the **Fungal Strains** section of the **Supplementary Information**. Of the total 59 fungal isolates, following phylogenetic characterization (by the Dettman lab), the isolates selected as the focus for the metabolomics analysis presented in this present chapter included 27 *Alternaria sect. A. alternata* isolates, 12 *Alternaria sect. A. arborescens* isolates, 16 *Alternaria sect. infectoriae* isolates - 5 isolates were excluded as an "outgroup" (as they were part of a larger project, beyond the scope of this thesis).

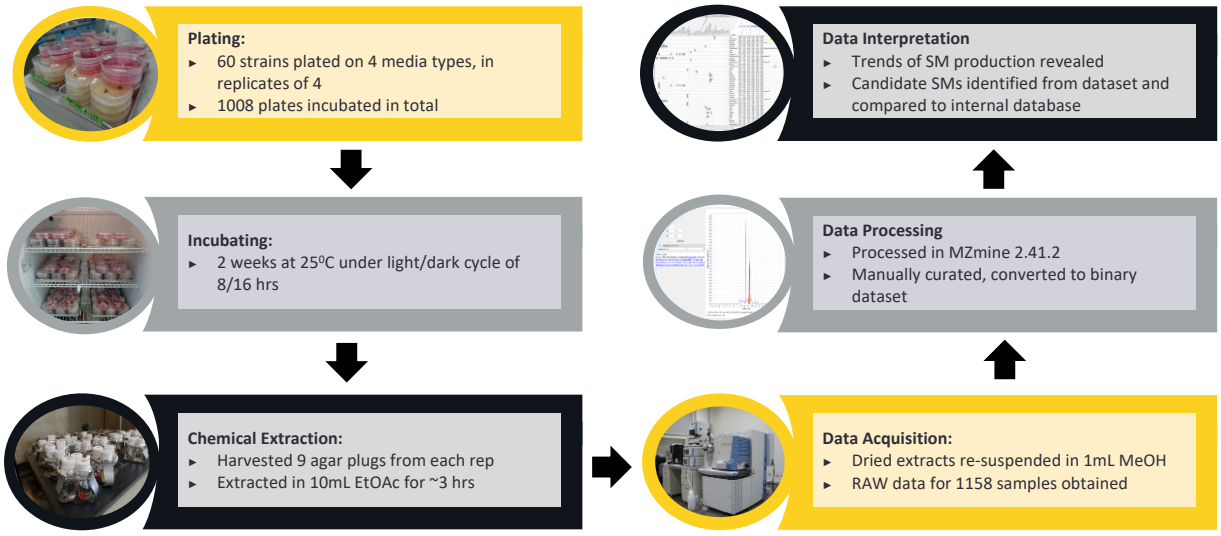

Figure 2.1. A generalized workflow for the methodology carried out for the experiment conducted in this Chapter of the present thesis. Each step is discussed in greater detail below.

2.2.3 Plating

The plating of the fungal samples was conducted over 3 batches of 20 samples (3x20 strains), whereby each batch was 20 strains, along with media control plates for each batch. Each batch was plated on different days, with no more than a week following the harvesting of one batch and the plating of the next batch. The fungal samples were plated in replicates of 16, whereby 4 replicates would be allocated to each of the utilized growth media conditions. The selection of growth media utilized included CYS80, DRYES, MMK2, and ZM2. A three-point inoculation approach was used, where three inoculation plugs were placed upon each plate, with the mycelium-side placed facing down, so as to be in direct contact with the culture medium.

2.2.4 Growth Conditions

After plating the fungal samples, they were stacked on trays (exemplified in the illustration next to "Plating" in **Figure 2.1**) and incubated under conditions of 25°C and a light/dark cycle of 8/16

hrs for 14 days. During the incubation period, trays were rotated relative to the light every 4 days.

2.2.5 Metabolite extractions

Nine agar plugs were harvested from each fungal replicate. The mycelium/agar plugs for each replicate were then stored in distinct scintillation vials and stored at -20℃ until they were able to undergo chemical extraction. Extractions were performed through the addition of 10mL ethyl acetate (EtOAc) to each scintillation vial. EtOAc was utilized for its biphasic actions where it can be used to extract both polar and non-polar compounds. Each scintillation vial was covered with aluminum foil and placed on a rotary shaker. Extractions were allowed to proceed for about 3 hours, under constant shaking of about 160RPM. After the 3-hour extraction period, shaking was stopped and the extracts were transferred to separate, new scintillation vials using a new Pasteur pipette for each replicate, being careful to leave the agar plugs in the original vial(s). After the successful transfer of the extract to the correct vial(s), the extracts were subjected to vacuum-concentration with a GeneVac (SP Scientific EZ-2 Elite). Once the extracts had completely dried-down, they were stored at -20℃ until being prepped for UPLC-HRMS analysis.

2.2.6 UPLC-HRMS analyses

Dried metabolite extracts were allowed to thaw completely following their removal from -20℃ (often taking < 1 hour per batch). After they had thawed, they were each re-suspended in 1mL of methanol (MeOH) and placed in HPLC vials (Thermo Scientific, Virtuoso Surestop 2mL amber vials, Langerwehe, Germany). The 960 metabolite extracts were analyzed on a Thermo Ultimate 3000 UPLC coupled to a Thermo LTQ Orbitrap XL high resolution mass spectrometer and a

Thermo Dionex Ulitmate 3000 Diode array detector (190-800nm). Chromatography was performed using a Phenomenex C_{18} Kinetex column (50mm x 2.1mm ID, 1.7µm) with a flow rate of 0.35mL/min over 10-minutes run, using a gradient of acetonitrile (+ 0.1% formic acid) and water (+ 0.1% formic acid). Specifically, the run started at 5% acetonitrile and increased to 95% acetonitrile by the 4.5-minute mark. The gradient would then be held at 95% acetonitrile until the 8-minute mark, at which point, the gradient returned to 5% acetonitrile by the 9-minute mark. A 5% acetonitrile gradient was then held to the 10-minute mark to equilibrate the column to starting conditions. The HRMS was operated in electro-spray ionization mode (ESI^+), monitoring a range of 100-2000*m/z*). The ESI used a capillary temp of 320℃, a sheath gas flow of 40, an auxiliary gas flow of 5, a sweep gas flow of 2, a source voltage of 4kV, a capillary voltage of 35V, and a tube lens of 100V. Through UPLC-HRMS analysis, 1158 raw data files were generated – denoting the metabolic profile data of all 960 fungal samples, 48 media controls, and 150 MeOH blanks.

2.2.7 Data pre-processing

The vendor (Windows DLL) peak picking algorithm was applied to the raw data, so as to pick out the peaks in the data and to prepare the data for MZmine processing by conversion to centroid format. Peak picking was performed through the use of the open-source software Proteowizard's msConvert (downloaded from: https://proteowizard.sourceforge.io/). After peak picking, the data was imported into MZmine 2.41.2 (downloaded from: http://mzmine.github.io), an open-source application that allows for the processing of large batches of data (Adusumilli and Mallick, 2017). Once the data had been successfully imported into MZmine, it was subjected to the stages of pre-processing (discussed in **Chapter 1**).

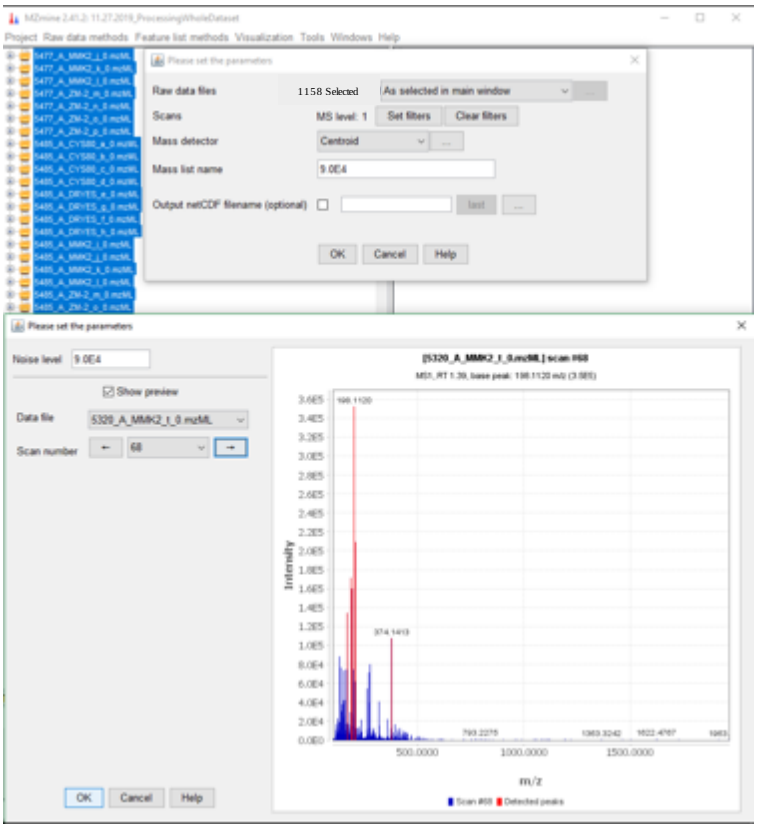

Figure 2.2. A screenshot of MZmine 2.41.2, in which, the centroid feature detection algorithm has been selected and the specified noise level parameter has been set to 9.0E4. Recognized *m/z* peaks are shown in red.

Feature detection is the first step in the metabolomic pre-processing pipeline (and is represented **Figure 2.2**, above). As the data imported into MZmine had been centroided prior to import, feature detection was carried out using the *centroid mass detection* algorithm. The *centroid mass detection* algorithm detects all points above the specified noise level as *m/z* peaks (Pluskal et al., 2010). A noise level of 9.0E4 was selected for the utilized dataset, based on inspection of the RAW data files/spectra.

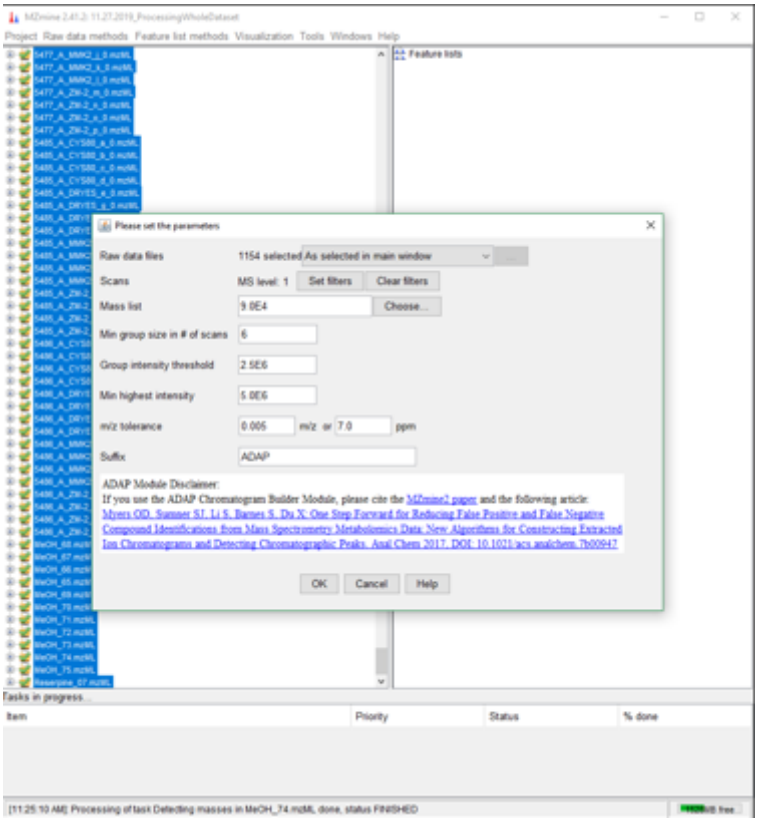

Figure 2.3. A screenshot of MZmine 2.41.2, in which, the ADAP chromatogram builder has been selected. The specified minimum group size in number of scans was set to 6, the group intensity threshold set to 2.5E6, the minimum highest intensity set to 5.0E6, and the *m/z* tolerance set to 7ppm (based on a caffeine standard in the extraction solvent).

The next step in the pre-processing pipeline is to apply the ADAP chromatogram builder algorithm to the dataset (represented by **Figure 2.3**). The parameters selected for the algorithm are shown in **Figure 2.3** but included specifying the minimum group size in number of scans as 6, the group intensity threshold to 2.5E6, the minimum highest intensity to 5.0E6, and the *m/z* tolerance set 7 ppm. The selection of these parameters was based on careful inspection of the raw data spectra, and the specified *m/z* tolerance was informed by observance of the standard feature included in the solvent.

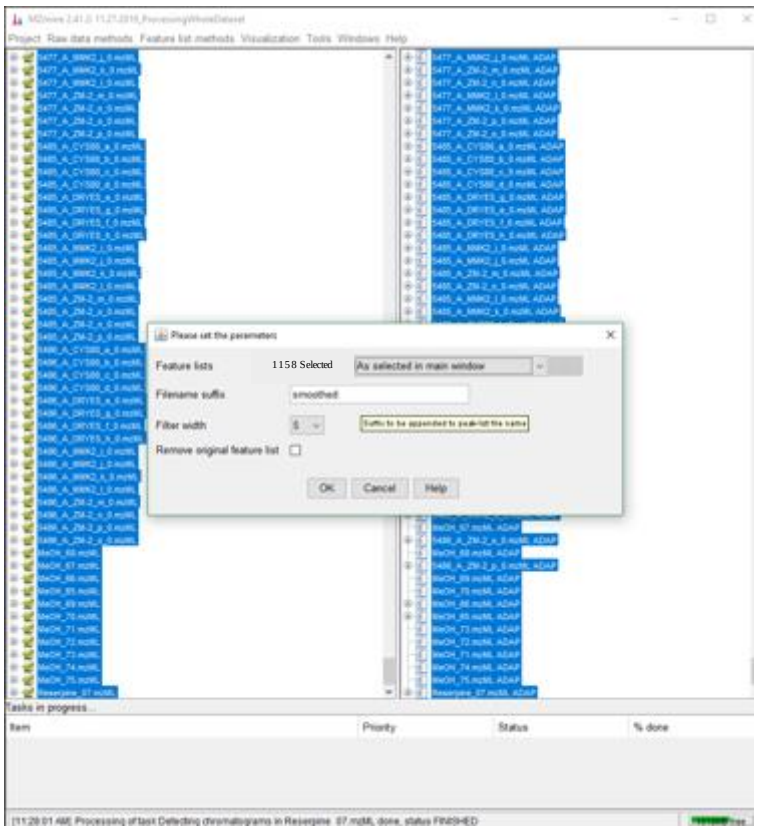

Figure 2.4. A screenshot of MZmine 2.41.2, in which, the smoothing algorithm has been selected. The specified filter width was 5.

The next step in the data pre-processing pipeline is the application of the smoothing algorithm to the dataset (represented by **Figure 2.4**). In order to apply the smoothing algorithm a dataset in MZMine, one needs only specify an appropriate filter width. The filter width utilized for the pre-processing the dataset generated in the present chapter, was 5.

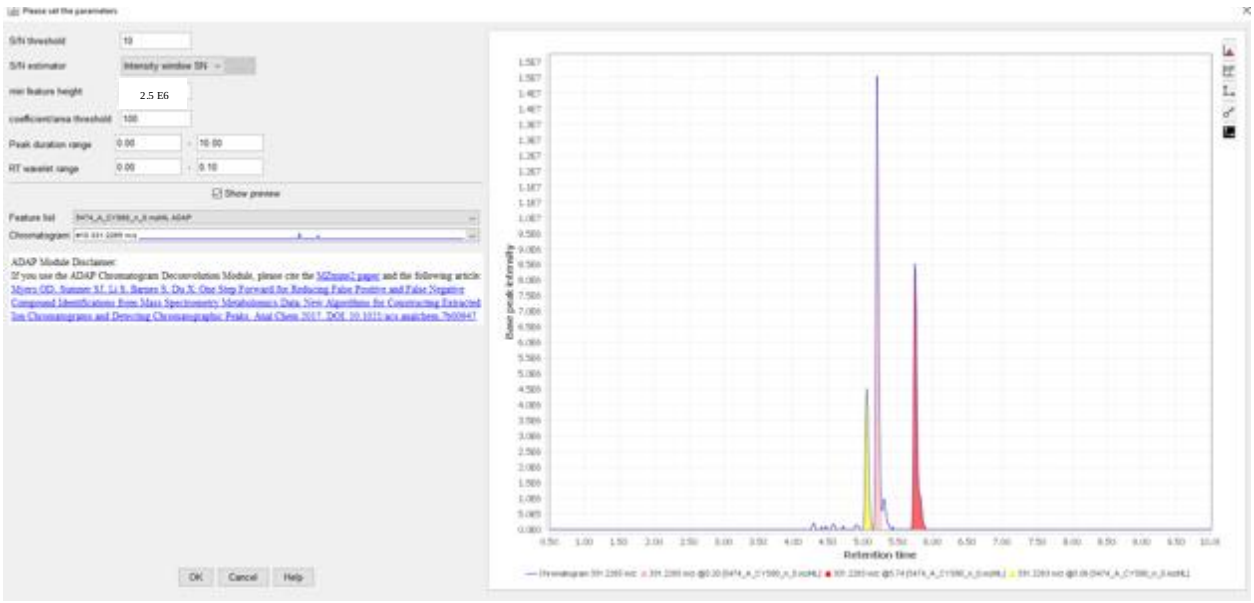

Figure 2.5. A screenshot of MZmine 2.41.2, in which, the chromatogram deconvolution was carried out (using a S/N threshold of 10, a minimum feature height of 2.5E6, a coefficient/area threshold of 100, the peak duration range set to capture the whole run, and the retention time wavelet range set to 0.10).

After smoothing the dataset, the next step in the data pre-processing pipeline is the deconvolution of each chromatogram object into its individual chromatographic peaks (Pluskal et al., 2010). The c*hromatogram deconvolution* algorithm is utilized to accomplish this task (and is shown in **Figure 2.5**). The parameters specified for the c*hromatogram deconvolution* algorithm applied to the dataset of the current chapter are presented in **Figure 2.5**.

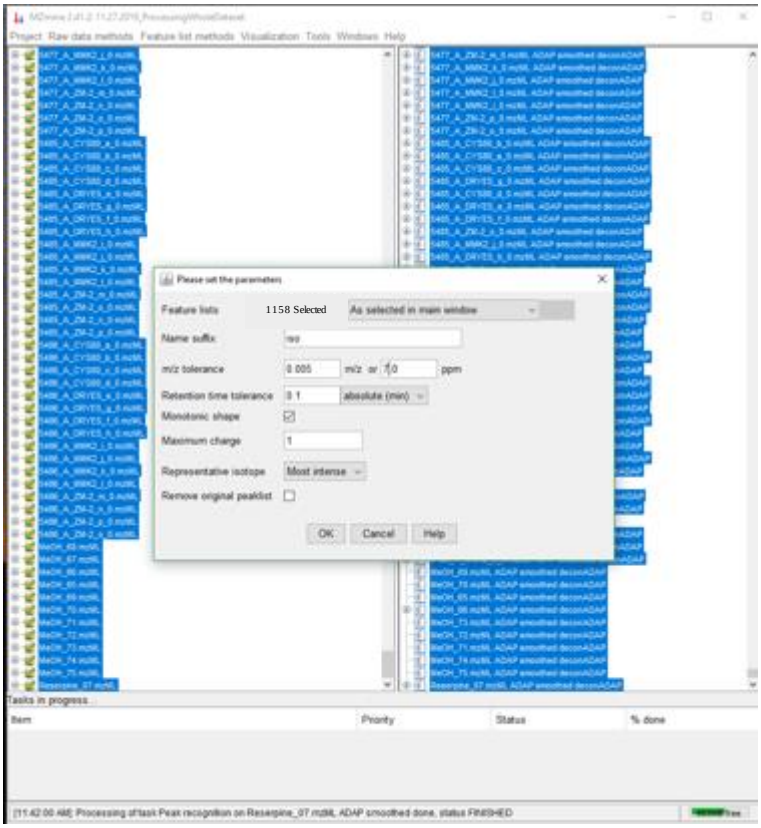

Figure 2.6. A screenshot of MZmine 2.41.2, in which, the isotope grouping was carried out **using a *m/z* tolerance of 7.0 ppm, a retention time tolerance of 0.1, the monotonic shape was selected, and a maximum charge of 1 was specified.** The representative isotope was set to be the most intense and the tailoring of the parameters was informed by the caffeine standard included in the extraction solvent.

Following chromatogram deconvolution, the next step in the data pre-processing pipeline is to apply the *isotope grouping algorithm* to the dataset. To apply the *isotope grouping algorithm*, the *m/z* tolerance, retention time tolerance, maximum charge, monotonic shape, and representative isotope parameters need to be specified. The values specified for these parameters are shown by **Figure 2.6,** above.

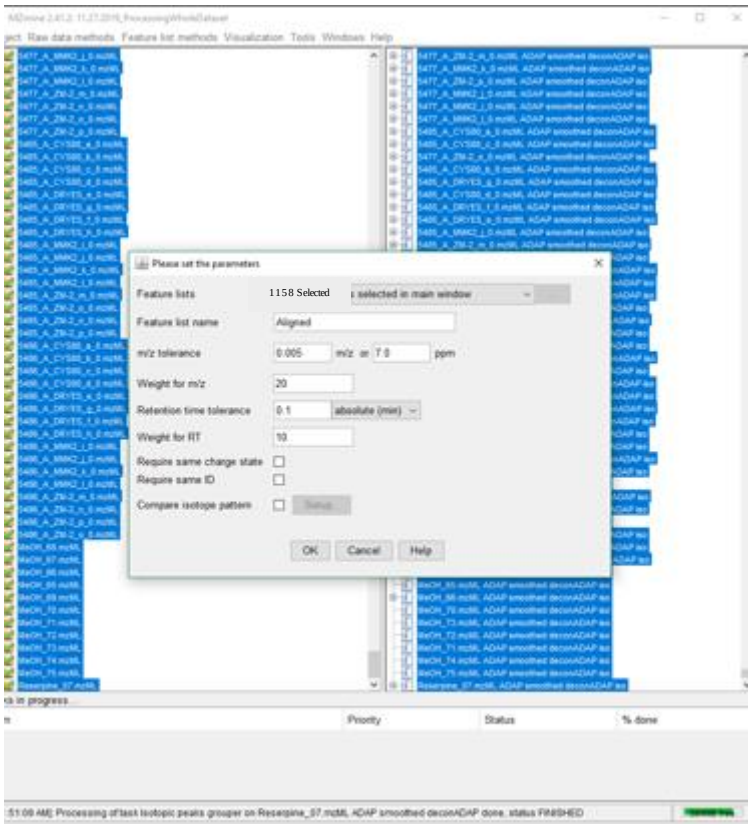

Figure 2.7. A screenshot of MZmine 2.41.2, in which, the join aligner was utilized to align peaks. The specified *m/z* tolerance was 7.0 ppm, a retention time tolerance of 0.1, the weight for *m/z* was set to 20, and the weight for RT was set to 10. As previously, the tailoring of the parameters was informed by the caffeine standard included in the extraction solvent.

After grouping identical peaks in the dataset according to their isotope pattern(s), applying the *join aligner* algorithm is the next step in the data pre-processing pipeline. The parameters specified for the utilized *join aligner* algorithm may be denoted by **Figure 2.7** but were as follows: *m/z* tolerance was set to 7.0 ppm, weight for *m/z* set to 20, retention time tolerance set to 0.1 min, and weight for RT set to 10. The selection of the utilized parameters was informed by inspection of the raw spectra, and accounting for the variance in the caffeine standard (included in the extraction solvent).

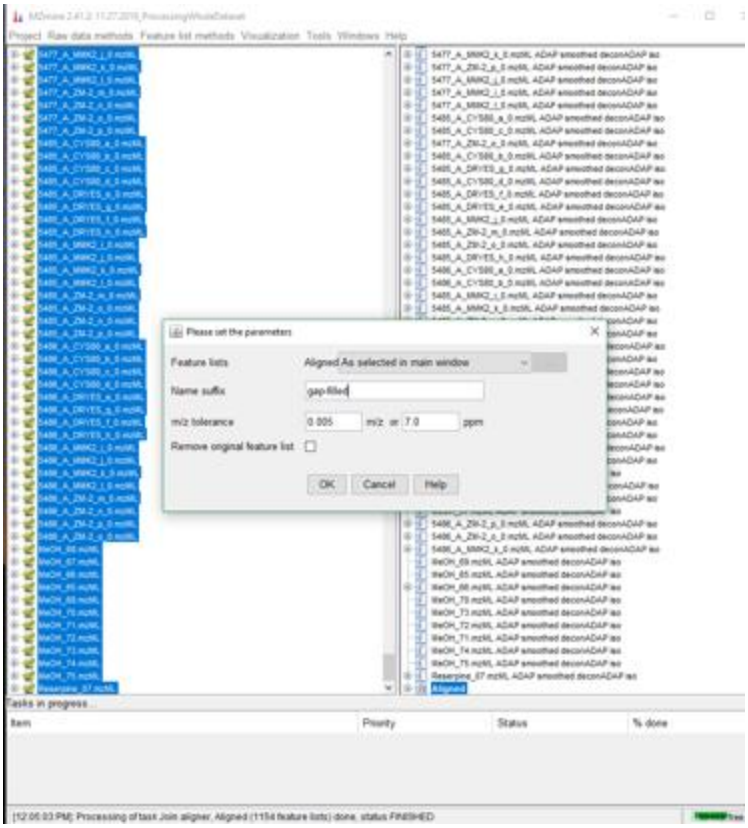

Figure 2.8. A screenshot of MZmine 2.41.2, in which, the same RT and *m/z* range algorithm was utilized for filling any gaps in the dataset. The specified *m/z* tolerance was 7.0 ppm, as was informed by the caffeine standard included in the extraction solvent.

After the application of the *join aligner* algorithm, the next step in the data pre-processing pipeline is the application of a *gap-filling* algorithm to the dataset. The *gap-filling* algorithm applied to the dataset generated in the current experiment was the "Same RT and m/z range gap filler", with the *m/z* tolerance parameter specified as 7.0 ppm (shown in **Figure 2.8**, above).

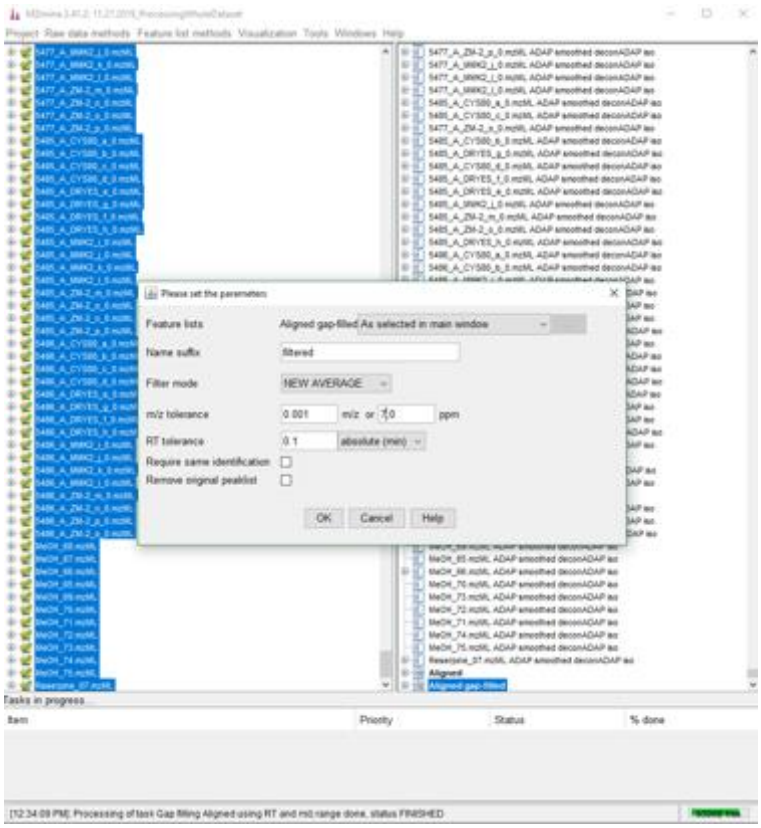

Figure 2.9. A screenshot of MZmine 2.41.2, in which, the duplicate peaks were removed from the dataset using the duplicate peak filtering algorithm. The filter mode was specified as "new average", and the *m/z* and RT tolerance range was informed by the caffeine standard included in the extraction solvent.

The final step in the data pre-processing pipeline is to remove duplicate peaks from the dataset. To accomplish this task, the *duplicate peak filtering* algorithm was utilized. The parameters selected for duplicate peak filtering may are detailed in **Figure 2.9**, above. After *duplicate peak filtering*, the pre-processed dataset was exported for curation. As such, it was exported as a matrix of mass feature values, with associated peak height data, separated by commas (a .csv file).

2.2.8 Data Processing

Exporting the processed dataset from MZmine, resulted in a data matrix of 1642 unique mass features. This dataset included media components and some "background noise" as represented in methanol blanks interspersed during the UPLC-HRMS analysis. Towards correcting the presence of these mass features, manual curation of the data matrix was required. In any instance where a mass feature was consistently detected above the noise level threshold in the media controls or methanol blanks, they were removed from the matrix, resulting in a reduction of mass features from 1642 features to 1343 unique mass features.

To convert the data matrix to a binary data matrix, the replicates of each strain were separated based on the media on which they were grown. If a given mass feature was detected in any 3 of the 4 replicates for that given media condition, a representative of the strain was created and the peak height of that mass feature, for that strain representative, was assigned a value of 1. If, however, the mass feature was not detected or was only detected in less than 3 of 4 of the replicates for that given medium, an associated peak height value of 0 assigned for that mass feature. In this way, the data matrix was reduced to a binary data matrix, whereby a 1 represented the production of a given mass feature and a 0 represented a lack of production of the given mass feature. Through this process, the associated representations of the culture extract replicates for each strain was reduced to a single representative. The data for each medium condition was parsed out accordingly, resulting in four separate binary data matrices (1 for each medium). Mass feature data from the four binary matrices were then summed to form a matrix of detection frequencies for each mass feature.

2.2.9 Statistical Analyses and Data Visualization

In order to perform statistical analyses on the data matrices (binary data matrix separated by medium, frequency of detection matrix, and non-binary data matrix separated by medium) the data was imported into an R environment (RStudio, available at: https://www.rstudio.com/products/rstudio/) and the MetaboAnalystR 3.0 package was utilized (retrievable at: https://github.com/xia-lab/MetaboAnalystR). The MetabolAnalystR package was utilized to normalize the non-binary matrices to an internal reference standard (the caffeine included in the extraction solvent), and to perform various multi- and uni-variate statistical analyses on the data. Specifically, MetaboAnalystR was utilized to apply a principal component analysis (PCA) and t-test to the data. MetaboAnalystR was also utilized to generate data visualizations, such as, the loadings plots, scores plots, heatmaps, dendrograms, and t-test results. The R coding necessary to accomplish all these analyses and visualizations may found in the **Supplementary Information**.

2.3 Results

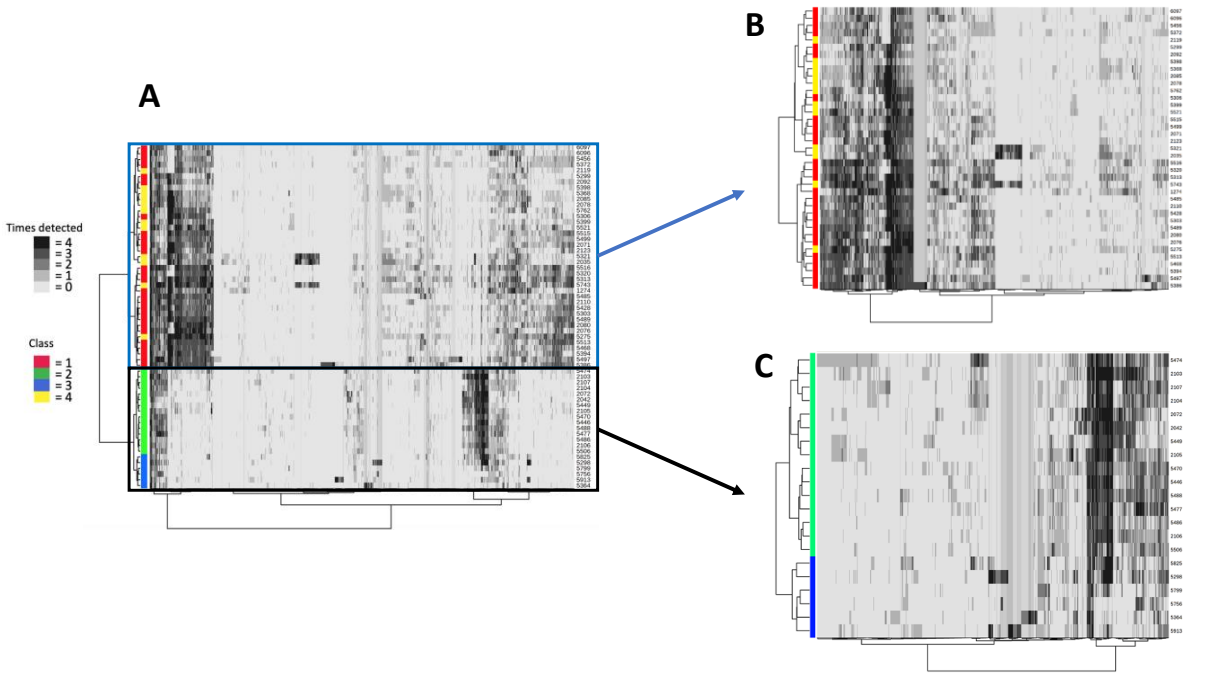

Figure 2.10. The chemical phenotypes of 59 isolates of *Alternaria* and 1 *Pithomyces* isolate, where heatmap values denote a consensus frequency data matrix of associated metabolite features (retention time + psuedomolecular ions). Heatmap A denotes the consensus chemical phenotype of the 60 studied isolates. Heatmap B denotes the consensus chemical phenotype of the studied isolates of Class 1 and 4 (the *Alternaria sect. A. alternata* (red) and *Alternaria sect. A. arborescens* group (yellow), respectively). **Heatmap C denotes the consensus chemical phenotype of the studied isolates of Class 2 and 3 (the** *Alternaria sect. A. infectoriae* (green) and the outgroup (blue), respectively). **Heatmap values in A, B, and C represent the frequency of detection for the mass features, over the four media conditions per isolate. The spectrum of values the heatmap values can have, are denoted in the legend to the left of heatmap A. Dendrograms on the left and bottom of each heatmaps were generated from hierarchical cluster analysis based of mass feature detection.**

The chemical phenotypes for each of the 60 studied isolates, on each of the different media conditions utilized, was generated in order to reveal patterns in mass feature production and metabolome diversity (**Figure 2.15 of Appendix A**). **Figure 2.15** shows the binary chemical phenotypes from each of the media conditions (**Figure 2.15A-D**). Owing to the difference in hierarchical clustering of the strains, and, therefore, SM production patterns amongst the media

conditions, it was evident that media conditions have a pronounced effect of SM production, that is maximized through using multiple media conditions. With the same strains exhibiting different SM mass feature profiles according to the medium for which they were grown on, identifying a trend from the profiles of each strain on each media would not be feasible. As such, it was necessary to use the consensus binary matrix to generate a consensus heatmap (**Figure 2.15A**). The consensus heatmap revealed a dichotomy in the dataset (**Figure 2.10A**).

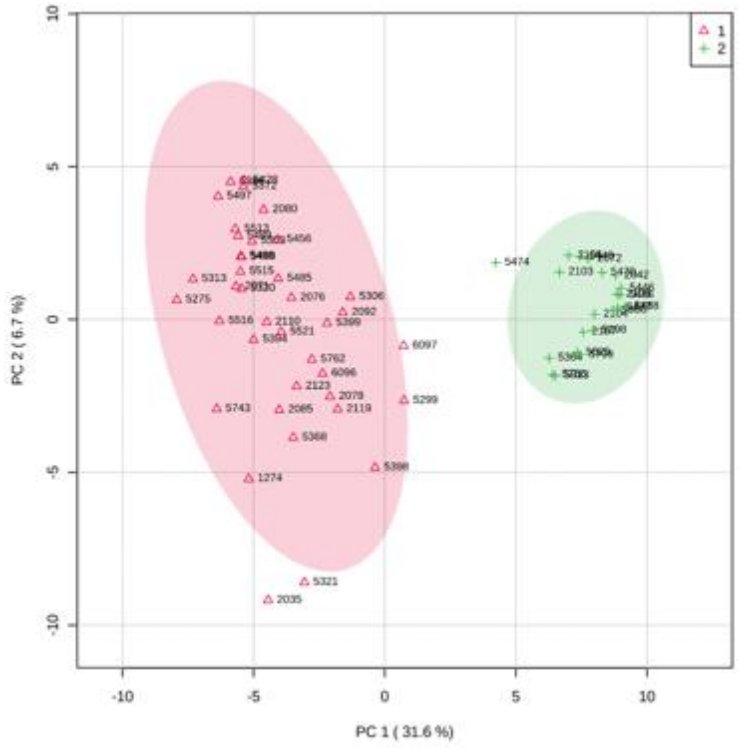

Figure 2.11. 2D scores plot illustrating 95% confidence ellipses for the binary data when the first two principal component (PC) score vectors are plotted against each other. Discrimination in the first component is between the group 1 (the *Alternaria sect. A. alternata* and *Alternaria sect. A. arborescens* group) and group 2 (the *Alternaria sect. infectoriae* and outgroup) samples. Group 1 may be denoted by the red triangles (△), and Group 2 by the green + symbols (+). The numbers next to each symbol represent the strain ID number. Separations along the second component indicate metabolic differences between the strains of each group to one another. The ellipses define the statistical significance of class separation and provide an illustration for the boundaries, which when crossed, reveal a clear separation of that/those samples from the rest of the group/data.

The consensus chemical phenotypes (**Figure 2.10A**) and PCA (**Figure 2.11**) utilize the mass-feature profiles to reveal a clear dichotomy in the dataset (further supported by the dendrogram represented by **Figure 2.16** of **Appendix A**). The observed dichotomy of the strains explained the majority of variance in both data models. This means that the data could be optimally separated into two distinct groups that could be analyzed separately from one another. The strains used in the current data models were classified using DNA barcoding genes (data not shown) to represent members of either the *Alternaria sect. A. alternata* or the *Alternaria sect. infectoria* species-complexes (**Figure 2.10B** and **Figure 2.10C**, respectively). Virtually all of the mass features detected in the *A. alternata* (and *A. arborescens*) of **Figure 2.10B** are completely absent from the SM profiles of any of the *A. infectoria* strains (and outgroup) (**Figure 2.10C**). The clear delineation observed therefore reveals that the SM profiles of the two groups are extremely different, resulting from fundamental differences in their SM production.

To reveal greater insights into differences in mass feature occurrence within species-complexes, group 1 (the strains selected from the *Alternaria sect. A. alternata* and the *Alternaria sect. A. arborescens*) was separated from group 2 (the strains selected from the *Alternaria sect. infectoriae* and outgroup), and for the remainder of the present work, they were referred to as the *A. alternata* group and the *A. infectoria* group, respectively.

In the PC1 vs PC2 scores plot (**Figure 2.11**), two strains from the *A. alternata* group, DET2035 and KAS5321, are observed to fall outside of the 95% confidence ellipse. This is indicative of potential differences in metabolite production from these strains when compared to the others of

the *A. alternata* group. As such, modelling the *A. alternata* group data separately from the *A. infectoria* group would allow further investigation into the mass features being utilized to distinguish DET2035 and KAS5321 from the other strains of the *A. alternata* group. Additionally, as the different fermentation media were shown to promote differential metabolome coverage (**Figure 2.15A, Figure 2.15B, Figure 2.15C, Figure 2.15D** of **Appendix A**), medium-dependent data modelling was performed to determine associations between unique strain-specific mass feature production and growth medium used for the *A. alternata* strains.

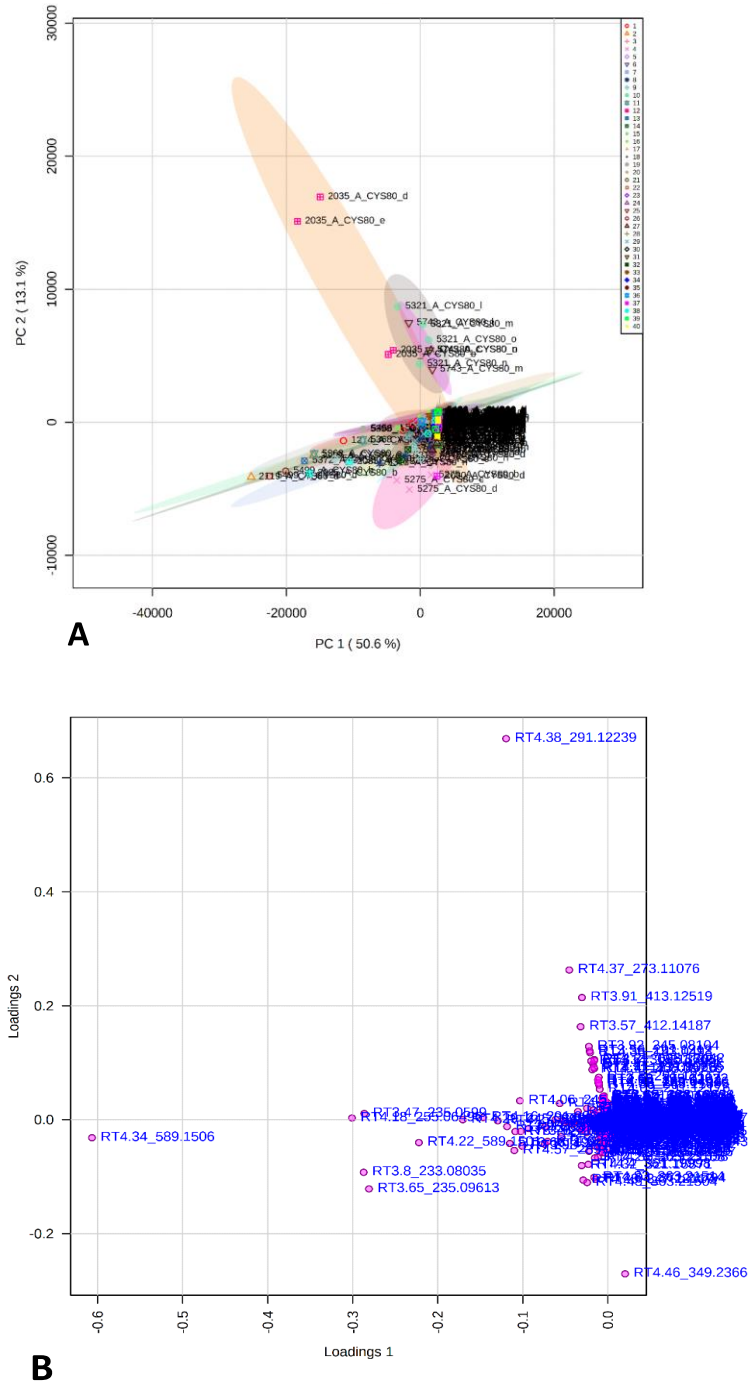

Figure 2.12. The score plot (A) and loadings plot (B) for the *Alternaria sect. A. alternata* group, grown on the CYS80 fermentation media. Both plots have PC1 on the x-axis and PC2 on the y-axis. Both plots **A** and **B** show the clustering of the *Alternaria sect. alternata* group strains, though plot **B** shows the mass features (retention time + pseudomolecular ion) that were utilized by the PCA clustering algorithm to cluster strains together.

The PCA resulting from the analysis of the members of the *alternata* group grown on CYS80 reveals that the PC1 can explain 50.6% of the variance in the data matrix, and PC2 can explain 13.1% (**Figure 2.12A**). Plotting PC1 against PC2 reveals a large cluster at the center of the cartesian plane, occupied by many strains of the group. Outside of that cluster, however, a unique cluster in the matrix is revealed that included all the replicates of DET2035, KAS5321 and KAS5743. Subsequent inspection of the loadings plot for PC1 against PC2 revealed that several key mass features are associated with the directional movement of these strains in the PCA model from the rest of the data matrix (**Figure 2.12B**). The key mass features produced by DET2035, KAS5321 and KAS5743 are highlighted in **Table 2.1**, below.

The PCA resulting from the analysis of the members of the *alternata* group grown on MMK2 revealed that PC1 and PC2 could explain 72.8% of the variance in the data (**Figure 2.17C** of **Appendix A**). As was observed in DRYES (data not presented), MMK2 did not promote as much diversity in SM production as did CYS80. This is suggested by the linear fit of PC1 to the bulk of the data. Inspection of the scores and loadings plot (**Figure 2.17A** and **Figure 2.17B** of **Appendix A**, respectively) revealed that the samples which cluster apart from PC1, represent all of the replicates of both DET2035 and KAS5321 (**Figure 2.17A**). As such, inspection of the loadings plot was necessary to determine the mass features responsible for distinguishing DET2035 and KAS5321 from all other samples in the MMK2 data (**Figure 2.17B**). The key mass features utilized by the PCA algorithm to distinguish DET2035 and KAS5321 as unique from the other samples are represented in **Table 2.1**.

The PCA resulting from the analysis of the members of the *alternata* group grown on ZM2 revealed that PC1 and PC2 could explain 70.7% of the variance in the data (**Figure 2.18A** of **Appendix A**). Inspection of the scores plot shows PC1 and PC2 fitting the data linearly in orthogonal directions (**Figure 2.18A**). Upon closer inspection, however, it was evident that there is a cluster at the intersection of PC1 and PC2 which encapsulates a large part of the data, but KAS6096 clusters along PC1, KAS1274 forms a cluster that bridges PC1 and PC2, KAS5513 clusters along PC2 and KAS5321, once again, clusters completely separate from all the other members of the *alternata* group (in the variable space) (**Figure 2.18A**). A comprehensive list of the features that drove the observed separation is not provided in the present work, but the separation of KAS5321 could be linked to the same mass features as were shown in the loadings plots of CYS80 and MMK2 (**Figure 2.12** and **Figure 2.17,** respectively). These key mass features may be found in **Table 2.1**.

The UPLC-HRMS run for DET2035 on CYS80 for a mass feature shown to be a major contributing factor to the separation of DET2035, KAS5321, and KAS5743 from the rest of the *alternata* group, is shown in **Figure 2.19** of **Appendix A**. Inspection of the MS spectrum reveals that the mass feature 291.12186 can also be linked to several adduct features (such as 273.11076, 245.11696, and several others that may be found in **Table 6.2** of the **Supplementary Information**). Inspection of the UPLC-HRMS runs for several of the other mass features which were shown to be a major contributor to the separation of DET2035, KAS5321, and KAS5743 revealed the presence of similar adducts.

Table 2.1. A list of the mass features which were predominantly responsible for the unique clustering patterns of DET2035, KAS5321, and KAS5743 when PC1 was plotted against PC2 to model the relationship between the members of the alternata group, when grown on the CYS80 and MMK2 fermentation media. These features distinguish DET2035, KAS5321, and KAS5743 from the main grouping on CYS80, KAS5321, and DET2035 from the main grouping on MMK2, and KAS5321 from the main grouping on ZM2.

Detected in (library strain ID)	Key mass features by parent pseudo-molecular ion (mass features within 7ppm of parent)				
	RT3.7_309.13205	**RT3.9_413.12619**	**RT4.15_323.14822**	**RT4.38_291.12239**	**RT4.47_293.13770**
DET2035, KAS5321, KAS5743	309.13205 ($[M+H]^+$) 273.11194 195.02859 123.04397	413.12619 ($[M+H]^+$) 435.10788 ($[M+Na]^+$) 273.11169 259.05958 245.08409 225.11174 195.02898 169.04939 123.08033	345.13031 ($[M+Na]^+$) 307.18974 ($[M-H_2O+H]^+$)	291.12239 ($[M+H]^+$) 273.11076 ($[M-H_2O+H]^+$) 245.11696 205.04916 195.02882 193.04930 169.04930 123.04361	293.13770 ($[M+H]^+$) 275.12708 ($[M-H_2O+H]^+$) 315.12071 ($[M+Na]^+$)

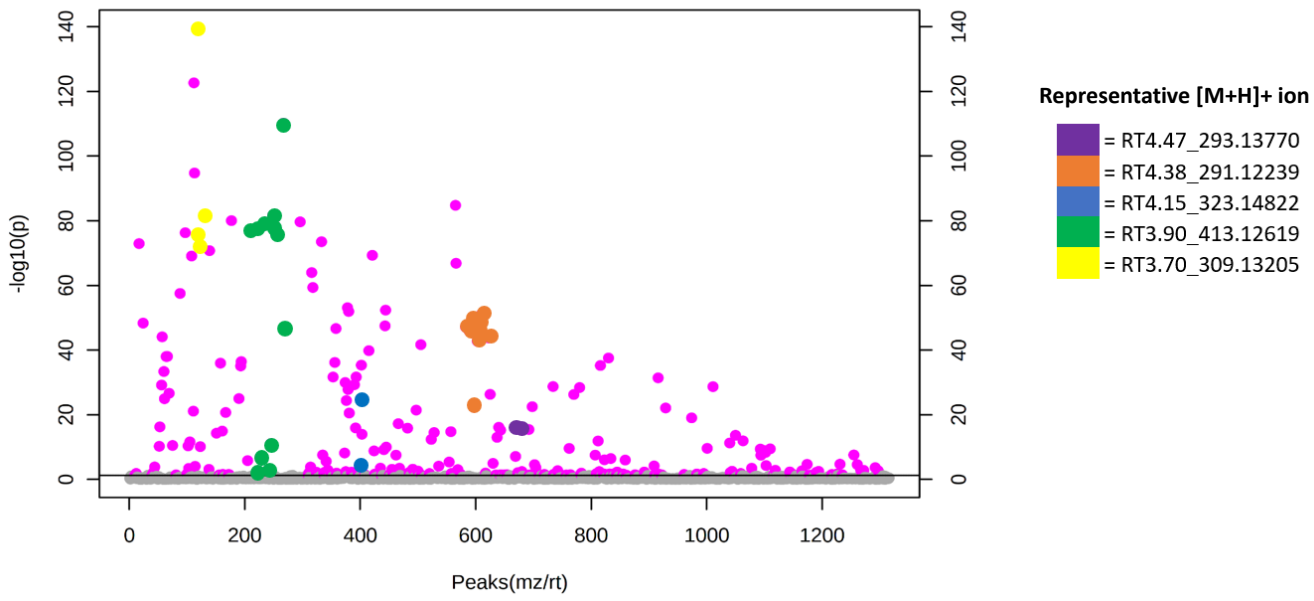

Figure 2.13. The T-test resulting from the alternata group on CYS80, with the features previously indicated to be the driving force in distinguishing DET2035, KAS5321, and KAS5743 highlighted in different colours (according to their representative [M+H]+ ion). The peaks are assigned along the x-axis in order of their retention time (for each mz/rt peak). The y-axis denotes the logarithmic p value of the associated peaks (and, therefore, directly proportional to the significance of the feature). Pink circles denote features produced by Group 2 (DET2035, KAS5321, and KAS5743), showing a statistically significant difference in production from Group 1 (all the other members of the *alternata* group). Colored circles denote

clusters of features which could be directly linked to one another (are adducts of the representative $[M+H]^+$ ion). Grey circles denote features bearing no statistical significance. $p < 0.05$.

The t-test revealed 182 features were produced by Group 2 (DET2035, KAS5321, and KAS5743) with statistically significant ($p < 0.05$) production patterns from Group 1 (all others of the *alternata* group) (**Figure 2.13**). The specific mass features are listed in **Table 2.1**, but their t-statistic, p-value, and false discovery rate are listed in **Table 6.2** of the **Supplementary Information**. The results of the t-test therefore indicate that the null hypothesis must be rejected. This means that there is statistical evidence to confirm the alternative hypothesis (that the strains comprising Group 2 are very different from the strains of Group 1, based on the profile of mass features they produce). Of the 182 features showing a statistically significant difference in production pattern, 26 (~15%) could be linked to the detection of fragments/adducts of a single family of compounds (see **Chapter 3**). As such, the results of the t-test suggest that the production of this family of compounds is a large part of what makes Group 2 unique.

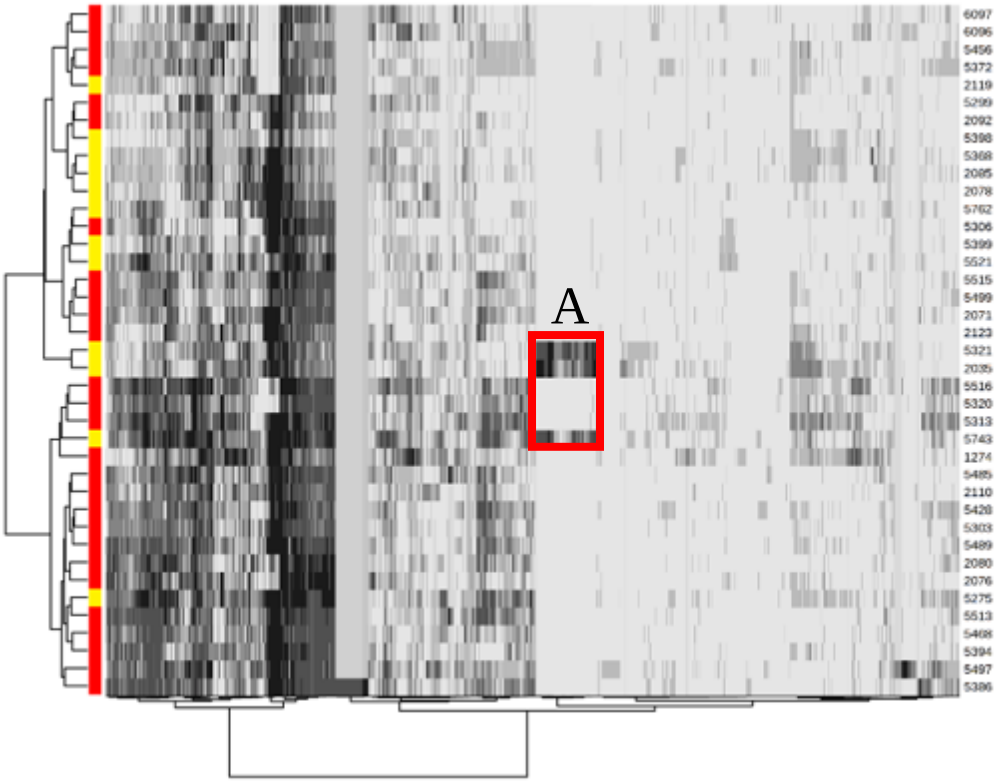

Figure 2.14. The consensus chemical phenotype of 39 *Alternaria* strains (representing the *alternata* group). The profile data was acquired on 4 different fermentation media and summed to generate a consensus frequency data matrix of associated metabolite features (retention time + psuedomolecular ions), produced through the combination of the binary matrices that resulted on each of the 4 different fermentation media. Associated metabolite features were grouped using hierarchical clustering analysis. **A** denotes mass features which stood out as being uniquely produced by DET2035, KAS5321, and KAS5743. Black pixels denote the presence of a metabolite feature (ie. 1), while the grey pixels denote the absence of the associated metabolite feature (ie. 0). *Alternaria sect. A. alternata* isolates are denoted by the red squares at the left of the heatmap and *Alternaria sect. A. arborescens* isolates by the yellow squares.

By visualizing the chemical phenotypes of the *alternata* group with a heatmap, the results of the PCA, scores plots and loadings plots results were corroborated. The heatmap revealed that the metabolite profiles of the *alternata* group consisted of core metabolome features and differential metabolome features (**Figure 2.14**). Differential metabolome features were those that were unique to one/few producers. Examples of such features are shown in **Figure 2.14A**. Core

metabolome features took the evident form of being shared/conserved across many/all of the members of the studied group. Based on the scores plots shown previously, it would be expected that the metabolite profile of KAS5321, DET2035 and KAS5743 would contain several unique features, which only they possess. **Figure 2.14A** reveals the presence of a just such a block of mass features. The characteristics of the specific mass features will be examined in detail in **Chapter 3** of the present thesis, though the time at which they eluted and their mass (*m/z* value) may be found in **Table 2.1**.

2.4 Discussion

The purpose of the current experiment was to utilize untargeted metabolomics to identify unique metabolite features in the SM profiles of 59 strains of *Alternaria*. In revealing unique metabolite features, it was hypothesized that this would lead to revealing unique biosynthetic gene clusters (BGCs).

The use of a variety of different media conditions was successful in promoting diverse metabolome coverage for the utilized *Alternaria* isolates as is evidenced by the profound effect that media composition had on metabolite production patterns. The most obvious examples could be the choice of fermentation media that promoted the production of several key mass features, which were utilized to distinguish KAS5321, KAS5743 and DET2035 from the others of the *alternata* group (CYS80 and MMK2). While as the SM profiles resulting from the fermentation of the *alternata* group on ZM2 allowed for KAS5321 to, again, be distinguished from the rest of the group, the features utilized to do so had not been reported in the other media types.

The mass features predominantly responsible for the clustering of DET2035 and KAS5321 (on CYS80 media) separately from the rest of the *alternata* group are listed in **Table 2.1**, but represent the detection of five different metabolites – some of which seemingly having common mass features and might belong to a molecular "family". Further confirmation of identification will be the focus of **Chapter 3** of the present thesis.

Historically, across *Alternaria*, the only active BGCs localized to ACs have been those responsible for producing HSTs. To this point, there has been a large amount of research interest

in the production of these HSTs as they often define the host range of the producing species. The majority of research on HSTs, however, was carried out prior to the advent of next-generation and long-read sequencing tools. As such, the localization of BGCs was based on electrophoretic karyotyping and visualized using a Southern blot probe (Witte et al., 2021). This means that in conducting discovery experiments, their approach had to be very narrow-focused and had to ignore the production of anything that could not be easily screened for. With the advent of high-quality long-read genome sequencing, coupled with high through-put untargeted metabolomics experiments, however, the focus can be much broader than just the HSTs.

The experimentation conducted for the present chapter is not without its limitations. Firstly, the use of more strains of *Alternaria* could have resulted in a more diverse/comprehensive SM profile than was observed in **Figure 2.14**. With a more diverse SM profile, perhaps more patterns of metabolite feature production would have stood out. The use of more strains, however, would cause for more signals and, therefore, noise in the dataset. As a result, it would have justifiably called for the pre-processing/processing pipeline to have been more stringent. Increased stringency could have caused for some features to get lost to the noise, and some could have been over-represented for their high abundance, thereby skewing results. As all of the metabolite features were quantified according to their relative abundance in the sample/instrument, one could argue that the data could already have been skewed by high abundance signals. While as normalization to a reference standard feature was performed prior to the use of any multi-variate statistics to curb any influence this could have, some influence is always an unavoidable reality of high-throughput metabolomics experimentation. Data processing and statistical analyses are

always going to cause a loss of some data. This explains why tens of thousands of signals were distilled down to 1642 features, which was then curated down to 1343 high-quality features.

In a metabolomics analysis, the goal is to select a model that optimizes the selection of high-quality data and minimizes the loss of low-abundance signals that may be of significance. The processing and statistical analyses selected for the exploration of the data aimed to minimize bias towards high abundance signal, partially through converting the dataset to a binomial/binary system. As a result, the complexity of the data would be minimized but, additionally, the relative abundance of a signal held no weight in the classification and clustering models applied. The metabolite feature was either detected, or it was not. Treating the data in this data processing pipeline could have led to the loss of certain low-abundance mass features, especially if the given mass features were detected in less than 75% of the replicates of a given strain on a given media. However, treating the data in this way allowed for the visualization of the chemical phenotypes of the studied isolates and the generation of a specific question to answer with statistics. As the results of the statistical analyses performed on the non-binary data matrices were generally in accordance with the chemical phenotypes generated with the binary data matrices, the selected pipeline for data processing was unlikely to have skewed data analysis/interpretation. It is unlikely that the detection of any of these five metabolites are a false positive as a result of the performed t-test analysis. The t-test analysis included false discovery rate (FDR) which calculates the likelihood that the given mass feature is a false positive. The associated FDR values for the mass features of the molecular family, were consistently less than 0.05, with some FDRs reaching 10^{-137}.

2.5 Conclusion

The purpose of the present chapter was to acquire a robust SM profile for a selection of *Alternaria spp.*, belonging to two major species-complexes (*Alternaria sect. A. alternata* and *Alternaria sect. infectoriae*), to assess for the detection/activation of unique BGCs. It was hypothesized that events of unique BGC activation would manifest as unique metabolite(s) detection in the dataset. After visualizing and interpreting the dataset in several ways (scores plots, loadings plots, t-test analysis, and heatmaps), just such a cluster of metabolites was revealed in three strains (DET2035, KAS5321, and KAS5743). Empirical proof could not be obtained to prove or disprove the hypothesis in the present chapter, though the obtained results could be utilized to support the hypothesis. Specifically, DET2035, KAS5743, and KAS5321 exemplified metabolite production patterns characteristic of the presence of at least one unique BGC shared between these strains. As such, strain DET2035 was selected for study in the chapters to follow, where the identity and structure of the mass features are elucidated.

2.6 References

Adusumilli, R., and Mallick, P. (2017). Data Conversion with ProteoWizard msConvert. *Methods Mol. Biol.* 1550, 339–368. doi:10.1007/978-1-4939-6747-6_23.

Akimitsu, K., Tsuge, T., Kodama, M., Yamamoto, M., and Otani, H. (2014). *Alternaria* host-selective toxins: determinant factors of plant disease. *J. Gen. Plant Pathol.* 80, 109–122. doi:10.1007/s10327-013-0498-7.

Andersen, B., Dongo, A., and Pryor, B. M. (2008). Secondary metabolite profiling of *Alternaria* dauci, A. porri, A. solani, and A. tomatophila. *Mycol. Res.* 112, 241–250. doi:https://doi.org/10.1016/j.mycres.2007.09.004.

Andersen, B., Krøger, E., and Roberts, R. G. (2002). Chemical and morphological segregation of *Alternaria* arborescens, A. Infectoria and A. Tenuissima species-groups. *Mycol. Res.* 106, 170–182. doi:10.1017/S0953756201005263.

Andersen, B., KrØger, E., and Roberts, R. G. (2001). Chemical and morphological segregation of *Alternaria* alternata, A. Gaisen and A. Longipes. *Mycol. Res.* 105, 291–299. doi:10.1017/S0953756201003446.

Andersen, B., Nielsen, K. F., Fernández Pinto, V., and Patriarca, A. (2015). Characterization of *Alternaria* strains from Argentinean blueberry, tomato, walnut and wheat. *Int. J. Food Microbiol.* doi:10.1016/j.ijfoodmicro.2014.11.029.

Andersen, B., Sørensen, J. L., Nielsen, K. F., van den Ende, B. G., and de Hoog, S. (2009). A polyphasic approach to the taxonomy of the *Alternaria* infectoria species–group. *Fungal Genet. Biol.* 46, 642–656. doi:https://doi.org/10.1016/j.fgb.2009.05.005.

Andersen, B., and Thrane, U. (1996). Differentiation of *Altemaria infectoria* and *Alternaria alternata* based on morphology, metabolite profiles, and cultural characteristics. *Can. J.*

Microbiol. 42, 685–689. doi:10.1139/m96-093.

Banks, J. G., Board, R. G., Carter, J., and Dodge, A. D. (1985). The cytotoxic and photodynamic inactivation of micro-organisms by Rose Bengal. *J. Appl. Bacteriol.* 58, 391–400. doi:10.1111/j.1365-2672.1985.tb01478.x.

Chaturvedi, V., Flynn, T., Niehaus, W. G., and Wong, B. (1996a). Stress tolerance and pathogenic potential of a mannitol mutant of Cryptococcus neoformans. *Microbiology* 142, 937–943. doi:10.1099/00221287-142-4-937.

Chaturvedi, V., Wong, B., and Newman, S. L. (1996b). Oxidative killing of Cryptococcus neoformans by human neutrophils. Evidence that fungal mannitol protects by scavenging reactive oxygen intermediates. *J. Immunol.* 156, 3836–3840.

da Cruz Cabral, L., Rodriguero, M., Stenglein, S., Fog Nielsen, K., and Patriarca, A. (2017). Characterization of small-spored *Alternaria* from Argentinean crops through a polyphasic approach. *Int. J. Food Microbiol.* 257, 206–215. doi:10.1016/j.ijfoodmicro.2017.06.026.

Frisvad, J. C., Andersen, B., and Thrane, U. (2008). The use of secondary metabolite profiling in chemotaxonomy of filamentous fungi. *Mycol. Res.* 112, 231–240. doi:https://doi.org/10.1016/j.mycres.2007.08.018.

Gacek, A., and Strauss, J. (2012). The chromatin code of fungal secondary metabolite gene clusters. *Appl. Microbiol. Biotechnol.* 95, 1389–1404. doi:10.1007/s00253-012-4208-8.

Greene, V., Barnett, L. H. L., Lilly, V., and Greene, B. (1953). The utilization of sugars by fungi. *West Virginia Agric. For. Exp. Stn. Bull.* 362T, 4–63. Available at: https://researchrepository.wvu.edu/wv_agricultural_and_forestry_experiment_station_bulletins/629 [Accessed June 30, 2021].

Hazen, E. L., and Reed, F. C. (1955). *Laboratory identification of pathogenic fungi simplified*.

Oxford, Blackwell Scientific Publications, Ltd.

Hocking, A. D., and Pitt, J. I. (1980). Dichloran-glycerol medium for enumeration of xerophilic fungi from low-moisture foods. *Appl. Environ. Microbiol.* 39, 488–492. doi:10.1128/aem.39.3.488-492.1980.

Hoffmeister, D., and NP, K. (2007). Natural products of filamentous fungi: enzymes, genes, and their regulation. *Nat. Prod. Rep.* 24, 393–416. doi:https://doi.org/10.1039/b603084j.

Hu, J., Chen, C., Peever, T., Dang, H., Lawrence, C., and Mitchell, T. (2012). Genomic characterization of the conditionally dispensable chromosome in *Alternaria arborescens* provides evidence for horizontal gene transfer. *BMC Genomics* 13, 171. doi:10.1186/1471-2164-13-171.

Jennings, D. B., Ehrenshaft, M., Pharr, D. M., and Williamson, J. D. (1998). Roles for mannitol and mannitol dehydrogenase in active oxygen-mediated plant defense. *Proc. Natl. Acad. Sci. U. S. A.* 95, 15129–15133. doi:10.1073/pnas.95.25.15129.

Kelman, M. J., Renaud, J. B., Seifert, K. A., Mack, J., Yeung, K. K.-C., and Sumarah, M. W. (2020). Chemotaxonomic Profiling of Canadian *Alternaria* Populations Using High-Resolution Mass Spectrometry. *Metabolites* 10, 238. doi:10.3390/metabo10060238.

King, A. D., Hocking, A. D., and Pitt, J. I. (1979). Dichloran-rose bengal medium for enumeration and isolation of molds from foods. *Appl. Environ. Microbiol.* 37, 959–964. doi:10.1128/aem.37.5.959-964.1979.

Lawrence, D. P., Gannibal, P. B., Peever, T. L., and Pryor, B. M. (2013). The sections of *Alternaria*: formalizing species-group concepts. *Mycologia* 105, 530–546. doi:10.3852/12-249.

Lawrence, D. P., Rotondo, F., and Gannibal, P. B. (2015). Biodiversity and taxonomy of the

pleomorphic genus *Alternaria*. *Mycol. Prog.* 15, 3. doi:10.1007/s11557-015-1144-x.

Meena, M., Prasad, V., Zehra, A., Gupta, V. K., and Upadhyay, R. S. (2015). Mannitol metabolism during pathogenic fungal-host interactions under stressed conditions. *Front. Microbiol.* 6. doi:10.3389/fmicb.2015.01019.

Nishimura, S., and Kohmoto, K. (1983). Host-specific toxins and chemical structures from *Alternaria* Species. *Annu. Rev. Phytopathol.* 21, 87–116. doi:10.1146/annurev.py.21.090183.000511.

Patriarca, A., da Cruz Cabral, L., Pavicich, M. A., Nielsen, K. F., and Andersen, B. (2019). Secondary metabolite profiles of small-spored *Alternaria* support the new phylogenetic organization of the genus. *Int. J. Food Microbiol.* 291, 135–143. doi:https://doi.org/10.1016/j.ijfoodmicro.2018.11.022.

Pluskal, T., Castillo, S., Villar-Briones, A., and Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11, 395. doi:10.1186/1471-2105-11-395.

Shoaib, A., Meraj, S., Nafisa, Khan, K. A., and Javaid, M. A. (2018). Influence of salinity and Fusarium oxysporum as the stress factors on morpho-physiological and yield attributes in onion. *Physiol. Mol. Biol. Plants* 24, 1093–1101. doi:10.1007/s12298-018-0570-z.

Simmons, E. G., and Roberts, R. G. (1993). *Alternaria* themes and variations (73). *Mycotaxon* 48, 109–140. Available at: https://www.cabi.org/isc/abstract/19942300558 [Accessed January 21, 2021].

Somma, S., Amatulli, M. T., Masiello, M., Moretti, A., and Logrieco, A. F. (2019). *Alternaria* species associated to wheat black point identified through a multilocus sequence approach. *Int. J. Food Microbiol.* 293, 34–43. doi:https://doi.org/10.1016/j.ijfoodmicro.2019.01.001.

Tsao, P. H. (1970). Selective Media for Isolation of Pathogenic Fungi. *Annu. Rev. Phytopathol.* 8, 157–186. doi:10.1146/annurev.py.08.090170.001105.

Wang, M., Fu, H., Shen, X.-X., Ruan, R., Rokas, A., and Li, H. (2019). Genomic features and evolution of the conditionally dispensable chromosome in the tangerine pathotype of *Alternaria* alternata. *Mol. Plant Pathol.* 20, 1425–1438. doi:https://doi.org/10.1111/mpp.12848.

Witte, T. E., Villeneuve, N., Boddy, C. N., and Overy, D. P. (2021). Accessory Chromosome-Acquired Secondary Metabolism in Plant Pathogenic Fungi: The Evolution of Biotrophs Into Host-Specific Pathogens . *Front. Microbiol.* 12, 700. Available at: https://www.frontiersin.org/article/10.3389/fmicb.2021.664276.

Witte, T., Harris, L., Nguyen, H., Hermans, A., Johnston, A., Sproule, A., et al. (2020). Apicidin biosynthesis is linked to accessory chromosomes in *Fusarium Poae* isolates. *BMC Genomics* Preprint. doi:10.21203/rs.3.rs-116075/v1.

Woudenberg, J. H. C., Groenewald, J. Z., Binder, M., and Crous, P. W. (2013). *Alternaria* redefined. *Stud. Mycol.* 75, 171–212. doi:10.3114/sim0015.

Woudenberg, J. H. C., Seidl, M. F., Groenewald, J. Z., de Vries, M., Stielow, J. B., Thomma, B. P. H. J., et al. (2015). *Alternaria* section *Alternaria*: Species, *formae speciales* or pathotypes? *Stud. Mycol.* 82, 1–21. doi:10.1016/j.simyco.2015.07.001.

Woudenberg, J. H. C., Truter, M., Groenewald, J. Z., and Crous, P. W. (2014). Large-spored *Alternaria* pathogens in section Porri disentangled. *Stud. Mycol.* 79, 1–47. doi:10.1016/j.simyco.2014.07.003.

Xie, H., Wang, X., van der Hooft, J. J. J., Medema, M. H., Chen, Z.-Y., Yue, X., et al. (2022). Fungi population metabolomics and molecular network study reveal novel biomarkers for

early detection of aflatoxigenic Aspergillus species. *J. Hazard. Mater.* 424, 127173. doi:https://doi.org/10.1016/j.jhazmat.2021.127173.

Yu, J. H., and Keller, N. (2005). Regulation of secondary metabolism in filamentous fungi. *Annu. Rev. Phytopathol.* 43, 437–458. doi:10.1146/annurev.phyto.43.040204.140214.

2.7 Appendix A

2.7.1 Chemicals and Solvents

Table 2.2. A list of the chemicals and solvents, along with their manufacturer's, utilized to perform the experimentation conducted in Chapter 2 of the present thesis.

Chemical	Manufacturer
Ethanol (EtOH)	Thermo-Fisher Scientific
Ethyl acetate (HPLC-grade)	EMD Millipore
Methanol (LCMS-grade)	VWR Chemicals
Acetonitrile (Optima, LCMS-grade)	Thermo-Fisher Scientific
Water (Optima, LCMS-grade)	Thermo-Fisher Scientific
Nanopure water	Thermo-Fisher Scientific, Barnstead Nanopure
Sucrose	Anachemia
Corn meal	Unico
Yeast extract	Oxoid
$MgSO_4 \cdot 7H_2O$	BDH
Agar	VWR
Dichloran	Sigma-Aldrich
Rose bengal	Sigma-Aldrich
Mannitol	Sigma-Aldrich
Murashige and Skoog Salts	Phytotechnology Laboratories
Molasses	Grandma's
Oatmeal	Quaker Oats Company
D-glucose	Sigma-Aldrich
$CaCO_3$	Thermo-Fisher Scientific
Edamin	Sigma-Aldrich
$(NH_4)_2SO_4$	Thermo-Fisher Scientific
Formic acid	Sigma-Aldrich

2.7.2 Media Formulations

2.7.3 Composition of the CYS80 medium

CYS80 medium was utilized as one of the growth media for the untargeted metabolomics experiment conducted in this chapter (on plates). It is comprised of the following:

Sucrose	80.0g
Corn meal	50.0g
Yeast extract	1.0g
$MgSO_4 \cdot 7H_2O$	0.5g
Agar	20.0g
Nanopure water	to 1.0L

2.7.4 Composition of the DRYES medium

DRYES medium was also utilized as one of the growth media for the untargeted metabolomics experiment conducted in this chapter (on plates). It is comprised of the following:

Sucrose	150.0g
Yeast extract	20.0g
$MgSO_4 \cdot 7H_2O$	0.5g
Dichloran (0.2% in EtOH)	1.0mL
Rose bengal (5%, w/v)	0.5mL
Agar	20.0g
Nanopure water	to 1.0L

2.7.5 Composition of the MMK2 medium

MMK2 medium was the third kind of growth media utilized for the untargeted metabolomics experiment conducted in this chapter (on plates). It is comprised of the following:

Mannitol	40.0g
Yeast extract	5.0g
Murashige and Skoog Salts	4.3g
Agar	20.0g
Nanopure water	to 1.0L

2.7.6 Composition of the ZM2 medium

ZM2 medium was the final kind of growth media utilized for the untargeted metabolomics experiment conducted in this chapter (on plates). It is comprised of the following:

Molasses	5.0g
Oatmeal	5.0g
Sucrose	4.0g

D-glucose	1.5g
$CaCO_3$	1.5mg
Edamin	0.5g
$(NH_4)_2SO_4$	0.5g
Agar	20.0g
Nanopure water	to 1.0L

2.7.7 Additional Results

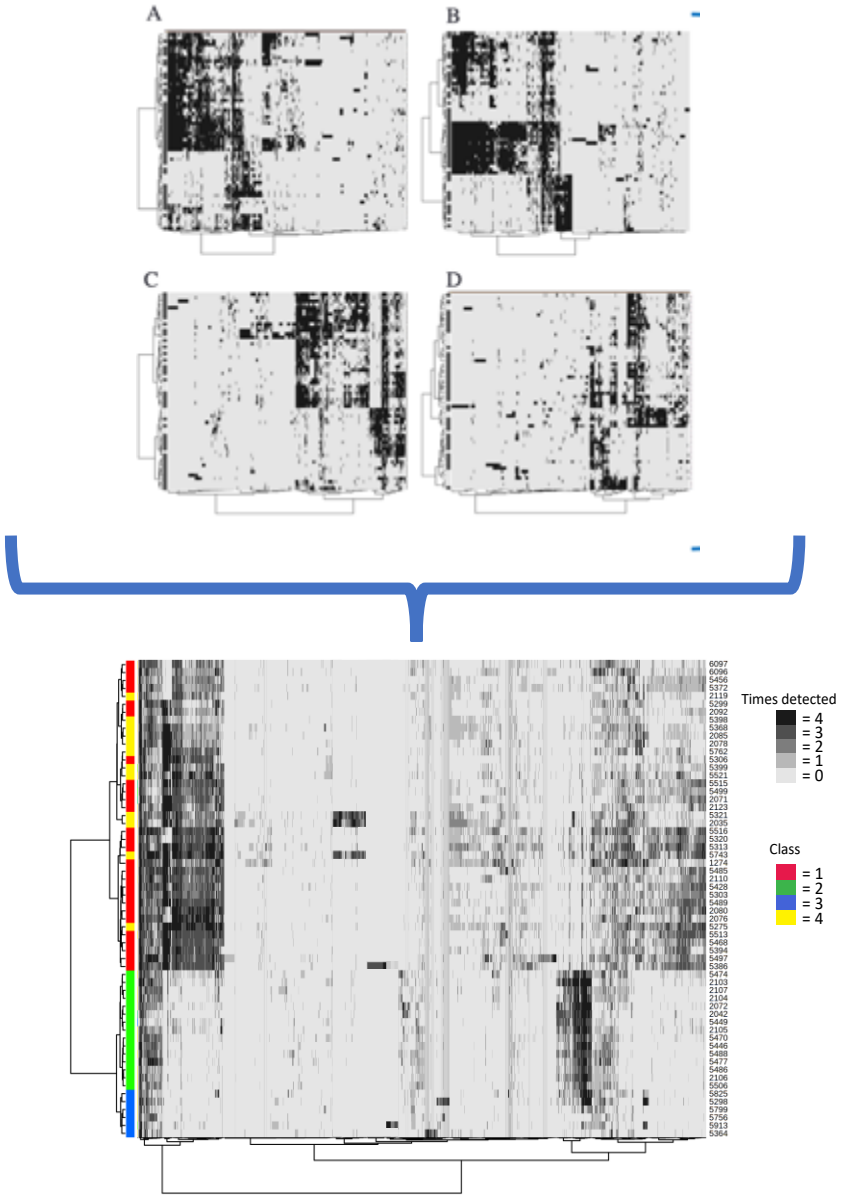

Figure 2.15. The chemical phenotypes of 59 isolates of *Alternaria* and 1 *Pithomyces* isolate. Heatmaps A, B, C, and D denote the chemical phenotype of all the isolates when grown on CYS80, DRYES, MMK2, and ZM2, respectively. Heatmap values denote a binary data matrix of associated metabolite features (retention time + psuedomolecular ions), which if detected in the 3 or more of the four replicates for the given media type, were assigned a 1 in the matrix (denoted by black squares in the heatmaps). Otherwise, the value for the given mass feature would be assigned a zero (denoted by light grey in the heatmaps). Heatmap E denotes the consensus chemical phenotype of the 60 studied isolates. Heatmap values in E represent the frequency of detection for the mass features, over the four media conditions per isolate. Dendrograms on the left and bottom of each heatmaps were generated from hierarchical cluster analysis based of mass feature detection. In heatmap E, the feature detection frequency may be denoted by the legend. Class 1 (red) represents the isolates belonging to the *Alternaria sect. A. alternata*, Class 2 (green) represents the isolates belonging to the *Alternaria sect. A. infectoriae*, Class 3 (blue) represents the "outgroup", and Class 4 (yellow) represents the isolates belonging to the *Alternaria sect. A. arborescens*.

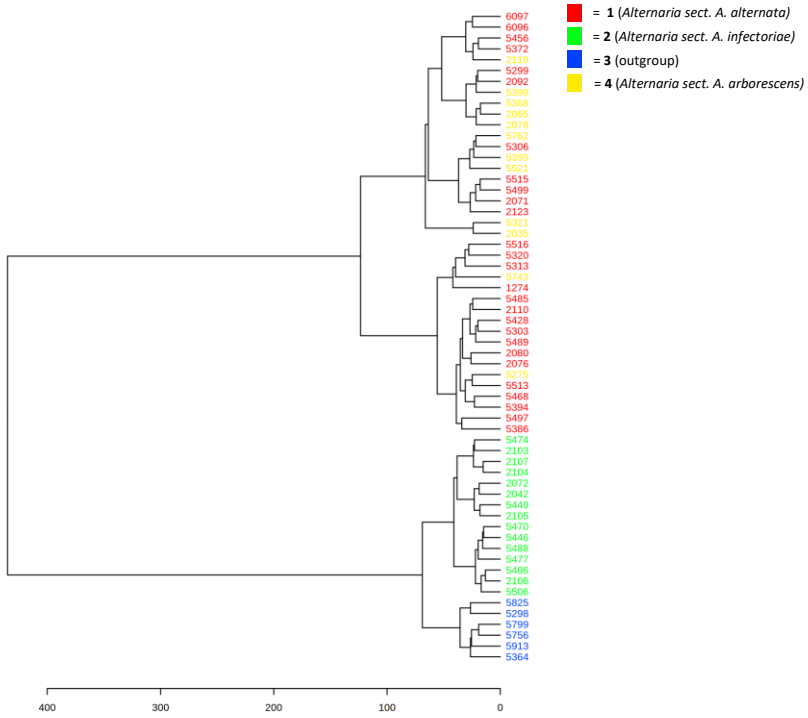

Figure 2.16. The dendrogram generated from the application of the Euclidean distance and Ward's Linkage hierarchical clustering algorithm to the consensus frequency data matrix of the 59 isolates of *Alternaria* and 1 *Pithomyces* isolate studied. The consensus frequency data matrix was produced using the frequency of detection for the mass features over the four media conditions per isolate. Class 1 (red) represents the isolates belonging to the *Alternaria sect. A. alternata*, Class 2 (green) represents the isolates belonging to the *Alternaria sect. A.*

infectoriae, Class 3 (blue) represents the "outgroup", and Class 4 (yellow) represents the isolates belonging to the *Alternaria sect. A. arborescens*.

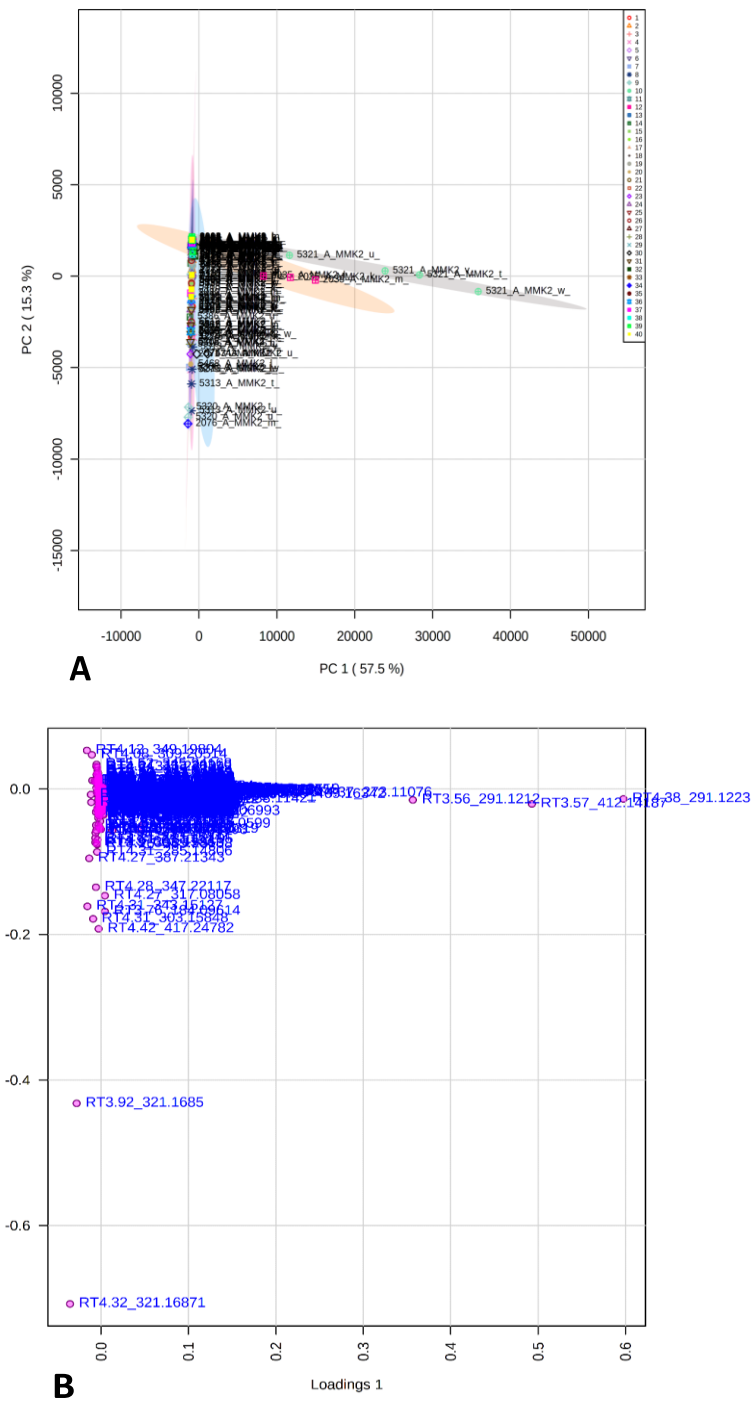

Figure 2.17. The score plot (A) and loadings plot (B) for the *Alternaria sect. A. alternata* group, grown on the MMK2 fermentation media. Both plots have PC1 on the x-axis and PC2 on the y-axis. Both plots **A** and **B** show the clustering of the *Alternaria sect. alternata* group strains, though plot **B** shows the mass features (retention time + pseudomolecular ion) that were utilized by the PCA clustering algorithm to cluster strains together.

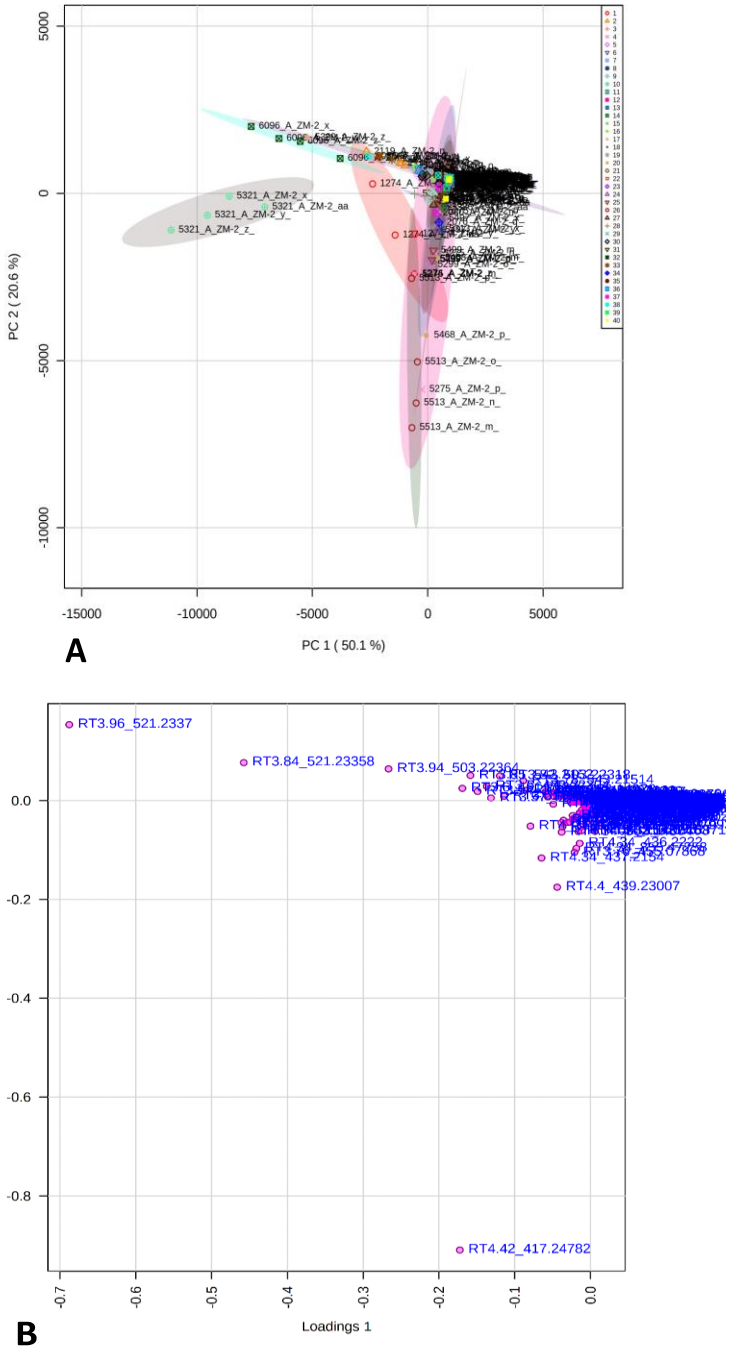

Figure 2.18. The score plot (A) and loadings plot (B) for the *Alternaria sect. A. alternata* group, grown on the ZM2 fermentation media. Both plots have PC1 on the x-axis and PC2 on the y-axis. Both plots **A** and **B** show the clustering of the *Alternaria sect. alternata* group strains, though plot **B** shows the mass features (retention time + pseudomolecular ion) that were utilized by the PCA clustering algorithm to cluster strains together.

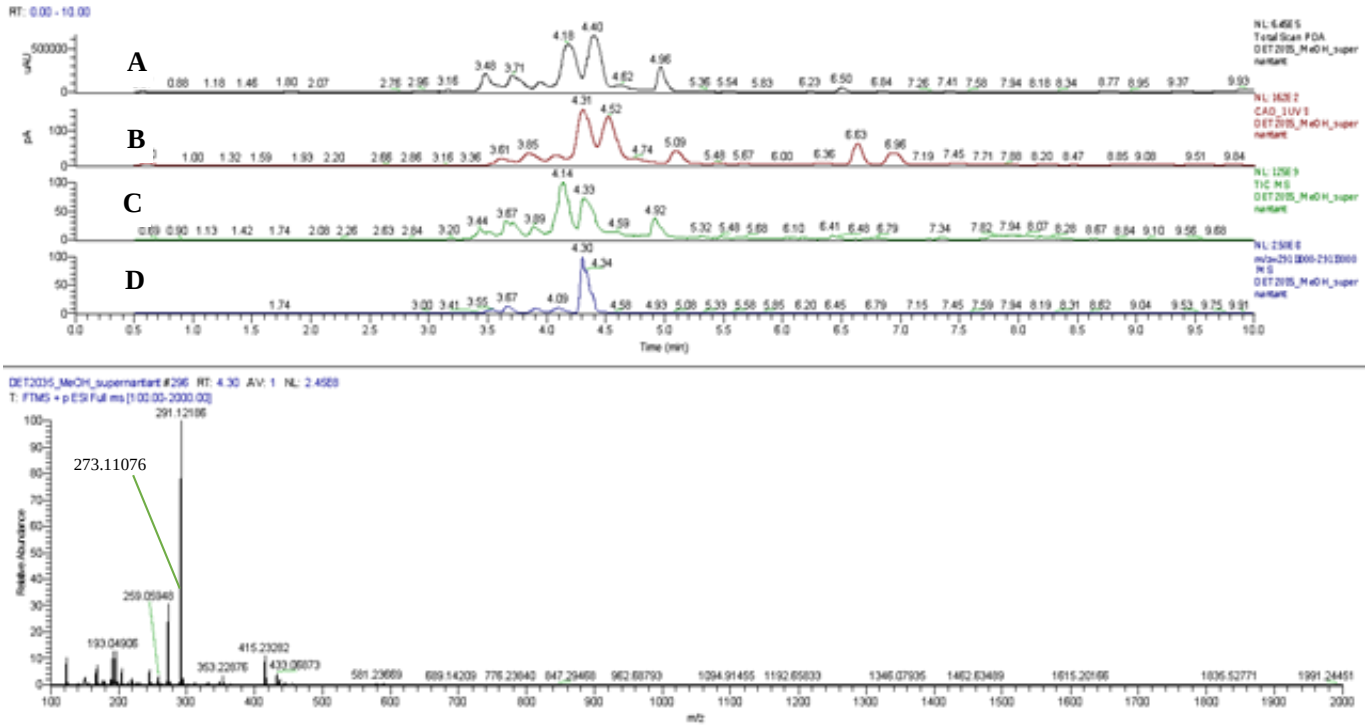

Figure 2.19. An example of the spectra obtained from a UPLC-HRMS run. The photo diode array (PDA) (**A**), charged aerosol detector (CAD) (**B**), total ion chromatograph (TIC) (**C**), extracted ion chromatogram (XIC) (**D**) and MS spectrum (**E**) were monitored for each run and exemplified herein.

Chapter 3

3 Candidate Natural Product Isolation, Purification, and Elucidation

3.1 Introduction

Alternaria, like many of the Dothideomycetes, are known for producing a vast diversity of SMs. With the large amount of interest they have garnered and their reputation as a pathogen to many integral crops worldwide, at least 270 metabolites from *Alternaria* fungi have been reported in the past few decades (Lou et al., 2013). Despite the efforts that have gone into discovering these metabolites, there is still much they produce that remains unknown. Furthermore, the synthesis, purpose and effects of many of those discovered still remains to be elucidated.

The toxic SMs that *Alternaria* produce are often classified according to the function they serve. The most commonly utilized terms are the virulence factors (VFs) and pathogenicity factors (PFs). The VFs are typically thought of encapsulating the nHSTs discussed in **Chapter 1** of the present thesis. In *Alternaria,* 30 nHSTs are known and characterized including alternariol (AOH), altenuene (ALT), tenuazonic acid (TeA), alternariol monomethyl ether (AME), tentoxin (TEN), curvularin, and altertoxin I, II, and II (Andersen et al., 2015; Lee et al., 2015; Meena et al., 2017). TEN has been found to inhibit chloroplast photo-phosphorylation by binding to ATP synthase and inhibiting ATP hydrolysis and synthesis, but the bioactivity of most of the nHSTs remains poorly understood (Steele et al., 1978). TeA, for instance, is one of the nHSTs most

frequently detected in *Alternaria spp*. isolates, yet, our understanding of its bioactivity is only that it has been implicated as inhibiting the process of protein synthesis in vivo (Meronuck et al., 1972). Curvularin is a nHST that has been reported in several *Alternaria spp*. and it has been implicated in arresting cell division by disrupting microtubule assembly (Kobayashi et al., 1988). How curvularin accomplishes this task, however, remains to be determined.

The PFs are typically thought of encapsulating the HSTs discussed in **Chapter 1** of the present thesis. The HSTs have predominantly been characterized in *A. alternata* pathotypes (or *f. sp*. epithets) to this point. HSTs have received much attention since their discovery by result of their interesting properties and specific activity. As discussed previously, the clusters of genes responsible for their biosynthesis have often been localized to ACs. The majority of research on HSTs in *Alternaria* was carried out prior to the advent of next-generation and long-red genome sequencing (Witte et al., 2021). Localization of the HST BGCs to ACs was typically based on electrophoretic karyotyping and visualized with a southern blot probe (Witte et al., 2021). Most of the known HST gene clusters have been characterized, however, like in the case of the AT- and ABR-toxin, there still remains some that have escaped characterization to this point. In order to determine the identity of the metabolites responsible for the unique clustering pattern of DET2035, KAS5321, and KAS5743, a targeted metabolomic experiment needs be embarked upon.

A traditional example of the use of targeted metabolomics can be found in its application for the targeted determination of trichothecenes and their derivatives in barley (Elosta et al., 2007). While as targeted metabolomics experiments have shown successful in confirming the

annotations made in an untargeted metabolomic analysis, its uses extend well beyond just that. For instance, a targeted metabolomics analysis was performed on pathogenic filamentous fungi whereby the high accuracy of ^{1}H-NMR spectroscopy were teamed with chemometric tools of metabolomics to try and differentiate three models of fungal strains from one another (Ząbek et al., 2017). Their results demonstrated that targeted metabolomic analyses could be utilized as a supporting taxonomical tool for some filamentous fungi (Ząbek et al., 2017).

Fungi have a highly complex chemical makeup, explaining why they are a major source of anti-infective agents and other medically-relevant compounds (Alhadrami et al., 2021). This complexity makes traditional blind chemical investigation a challenging process (Alhadrami et al., 2021). Using cheminformatics tools from metabolomics, the complexity of the data can be greatly reduced and, as a result, the most probable bioactive candidates can be selected with greater ease and confidence. Further, if those candidates are mined from an untargeted metabolomic analysis, a subsequent targeted metabolomic analysis can optimize the production and, therefore, yield of those select natural products. Untargeted metabolomic analyses have been utilized to reveal VFs in *Xylella fastidiosa* and *Pseudomonas aeruginosa* (Allain et al., 2019; Feitosa-Junior et al., 2019; Castro-Moretti et al., 2020; Depke et al., 2020). As the field of metabolomics is currently in its infancy, the guided approach provided by the use of untargeted and targeted metabolomic analysis for natural product discovery and guided isolation/elucidation is rare. An example of the successful utilization of such techniques can be found in the characterization of the apicidins in *F. poae* isolates (Witte et al., 2020).

The purpose of this study was to perform a scale-up growth fermentation of isolates shown to have produced several unique mass features in the work of the previous chapter. The goal was then to isolate, purify, and elucidate the identity of the metabolites corresponding to those mass features. It was hypothesized that the cluster of unique mass features discovered in **Chapter 2** represents the detection of a related molecular family of analogs sharing structural similarities.

3.2 Materials and Methods

3.2.1 Media, Chemicals and Solvents

The CYS80 medium was implicated in the previous chapter for encouraging the maximal production of the candidate mass features in DET2035. As such, 300 plates were made with the same composition as can be found in **Appendix A** of **Chapter 2**. The chemical, solvents, and materials utilized for the experiment of the current chapter may be found in **Appendix B** of the present Chapter.

3.2.2 Fungal Strains

The DET2035 strain was selected for the experimentation of the present chapter. It was obtained from the Canadian Collection of Fungal Cultures (DAOMC) of Agriculture and Agri-Food Canada (AAFC/AAC), in which, the selected strains belonged to the DET (Dr. J. Dettman) collection library.

3.2.3 Plating

Based on the results of the previous chapter, DET2035 was plated in 2 batches of 150 plates on CYS80 growth media, over the course of 2 days. The same protocol as the previous chapter was adhered to for the present experiment, whereby, 3 inoculation points were placed upon each plate with the hyphae placed facing down (so as to directly interact with the media).

3.2.4 Growth Conditions

Inoculated plates were incubated at 25^{o}C with a light/dark cycle of 8/16 hrs for 14 days. The plate trays were rotated relative to the light every 4 days.

3.2.5 Metabolite extractions

All of the contents of the cultures were harvested for extraction. The extractions were performed by first cutting up the cultures into small squares (hereafter referred to as fungal squares). In this way, the surface area of fungi interacting with the extraction solvent would be maximized. The fungal squares were then placed in a 2 L Erlenmeyer flask, until there was about 600 mL of fungal squares in each flask. Each harvest usually required the use of 5 flasks. ~1300 mL of ethyl acetate (EtOAc) was then added to each flask and covered with aluminum foil (so as to prevent the volatile solvent from escaping the vial). Extractions were allowed to proceed for a minimum of 3 hours, under constant shaking of about 160 – 175 RPM (on a rotary shaker). After the extraction period, shaking was stopped and the extracts were vacuum filtered into a new flask through use of a Buchner funnel using Whatman filter papers (Whatman 5230-110, Grade 230, diam 100 mm). The filtered extract was transferred to a round-bottomed flask and was concentrated in vacuo using a rotary evaporator (IKA RV10). The flask containing the extracted fungal squares, was then subjected to another round of extraction, whereby, ~1300 mL of ethyl acetate (EtOAc) was again added to each flask. After extraction, the second extraction was combined with the residue from the first extraction and concentrated to dryness in vacuo using a rotary evaporator.

A de-fatting step was performed by re-suspending the extract in methanol and the subsequent addition of the extract to a 2 L seperatory funnel. An equal volume of hexanes was added, and the vessel was shaken vigorously. An emulsion formed and, as a result, the mixture needed be left for about an hour without agitation. The bottom and top layer were then separated (once the boundaries of each phase had become very clear) and each collected in a respective round-

bottomed flask. The bottom layer, which was the methanol fraction, was collected, and a 500 μL aliquot of the extract was placed in a HPLC vial (for submission for UPLC-HRMS analysis). The remainder of the methanol fraction was concentrated in vacuo using a GeneVac EZ-Elite 2 (SP Scientific, Ipswich, UK). The resulting sample was stored at -20°C until the results of the UPLC-HRMS could confirm that the feature(s) of interest were indeed produced in sufficient quantity.

3.2.6 Target Metabolite Identification

The bulk extract of DET2035 was analyzed by UPLC-HRMS (Waters Ultimate 3000 and Thermo-Fisher Scientific LTQ Orbitrap XL). The protocol carried out was the same as was utilized in **Chapter 2**. This was necessary as it was under those conditions that the mass features were first discovered. Extracted ion chromatograms enabled identification of features of interest and their relative quantification. The extracted ion chromatograms were generated through subjecting the RAW data to mass range filtering in Xcalibur (Thermo-Fisher Scientific).

Additionally, in-silico MS/MS analysis was run around the window of time that the mass features elute from column (from 3.5 to 5.5 min). This allowed for the in-silico fragmentation pattern of the features of interest to be examined, and their identities to be more closely elucidated. MSFinder 3.44 (obtained from: http://prime.psc.riken.jp/compms/msfinder/main.html) was utilized to interpret the fragmentation pattern of the features of interest and generate a list of potential candidate structures. It does so by comparing the imported spectra to those that it can predict for the compounds available from all the databases which it has access to (such as PubChem, ChemSpider, KEGG, MetaCyc, MINE, CHEBI, UNPD, etc.).

The suspected identity of the features was also predicted through comparison of the pseudomolecular ion masses to that of an internally-compiled database of the 300+ SMs which have been identified across literature in *Alternaria* species. The comparison of the list of candidate features to the internally-complied database was automated using an R-script. If the mass features were within 5 ppm of a $[M+H]^+$, $[M+Na]^+$, or $[M-H_2O]^+$ in the database, the name of that compound would be outputted as a match to the respective mass feature. The script was written by Thomas Witte of the Metabolomics laboratories of Dr. David Overy (of AAFC/AAC).

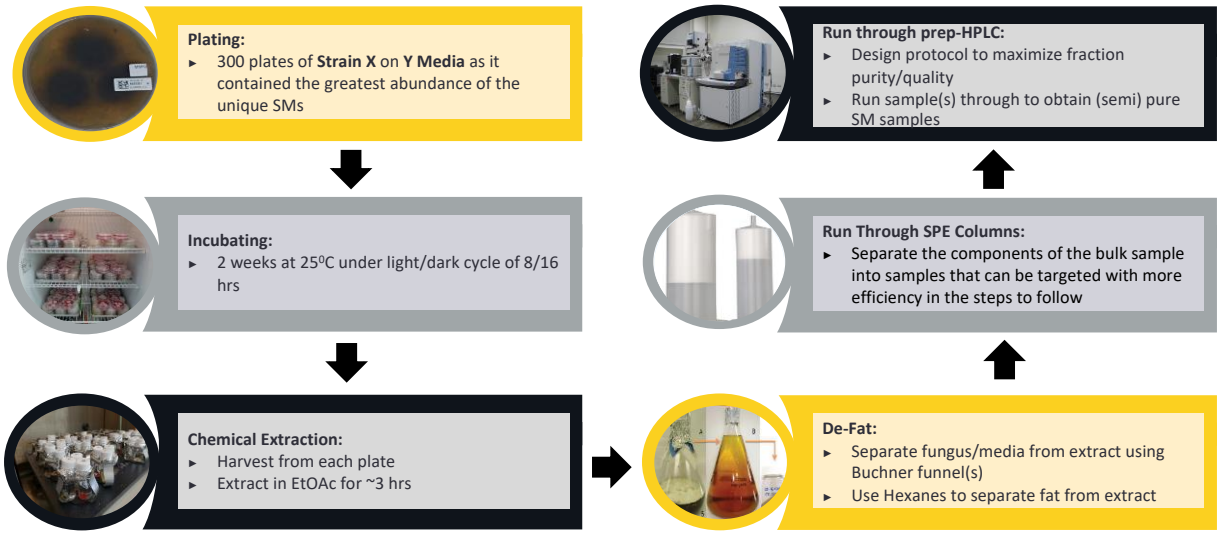

Figure 3.1. A workflow depicting the general procedure that will be employed for the isolation of the target features from the respective strains. Each step is discussed in greater detail throughout this materials and methods section.

3.2.7 Targeted Metabolite Isolation

The dried bulk extract was re-suspended in as little MeOH as was necessary to completely dissolve the sample (about ~10 mL for the DET2035 bulk extract). It was then transferred into a new pre-weighed scintillation vial through a .45 μm PFTE Acrodisc (Pall Corporation, MI,

USA) syringe filter to remove particulates. The filter was rinsed with methanol several times, and that methanol collected in the extract sample. The pre- and post-filtered bulk extracts are shown in **Figure 3.2,** below. After filtering out the particulates, the bulk extract was ready for HPLC fractionation.

Figure 3.2. The DET2035 bulk-extract sample, before (left) and after (right) filtering

3.2.8 Prep-HPLC

Fractionation of the bulk-extract was performed using a prep-HPLC (1260 Infinite Series, Agilent Technologies, CA, USA) and a Phenomenex Kinetex C18 100Å LC column (100 x 30.0 mm). In an attempt to cause the features of interest to elute at a similar time point to was observed in the UPLC-HRMS data, a gradient was chosen that closely resembled that utilized in the UPLC-HRMS run (discussed in previous chapter), but did not involve a steep ramp-up to 95% ACN. As such, the flow rate was 20.0 mL/min, and analysis was performed over an 11-minute run, using a gradient of ACN (+0.05% formic acid) and H_2O (+0.05% formic acid). The run started at 50% ACN and increased to 95% ACN by 4.5 minutes. 95% ACN was then held steady at 95% until the 8.0-minute mark, at which point, it was set to decrease to 5% ACN by the

10-minute mark, where it was held for a minute (until the 11-minute mark). After the 11-minute mark, the fraction collector ceased, and the column was allowed 2 minutes to equilibrate to 50% ACN. 500uL of bulk-extract was injected for each run and, as such, it took a total of 14 injections to fully fractionate the bulk-extract.

The fractions generated through prep-HPLC were then re-suspended in MeOH (in HPLC-vials), such that they all had a concentration of 500 μg/mL. The fractions were then placed into -20℃ until they could undergo UPLC-HRMS analysis.

3.2.9 UPLC-HRMS and MS/MS Analysis

The fractions were analyzed on the same Thermo Ultimate 3000 UPLC coupled to a Thermo LTQ Orbitrap XL high resolution mass spectrometer and a Thermo Dionex Ulitmate 3000 Diode array detector (190-800 nm) as was utilized in the previous chapter. The column utilized for chromatography was a Phenomenex C_{18} Kinetex column (50mm x 2.1 mm ID, 1.7 μm), and the flow rate was 0.35 mL/min. Analysis was performed over a 10-minute run, using a gradient of acetonitrile (ACN) (+ 0.1% formic acid) and water (+ 0.1% formic acid). Specifically, the run started at 5% ACN and increased to 95% ACN by the 4.5-minute mark. The gradient would then be held at 95% ACN until the 8-minute mark, at which point, the gradient returned to 5% ACN by the 9-minute mark. A 5% ACN gradient was then held to the 10-minute mark to equilibrate the column to starting conditions. The HRMS was operated in electro-spray ionization mode (ESI^+), monitoring a range of 100-2000 *m/z*). The ESI used a capillary temp of 320℃, a sheath gas flow of 40, an auxiliary gas flow of 5, a sweep gas flow of 2, a source voltage of 4 kV, a capillary voltage of 35 V, and a tube lens of 100 V.

MS^n fragmentation was performed on several extracts from the previous chapter, containing the mass features of interest (annotated ions) to the present chapter. Specifically, select ions detected in extracts from KAS5321 and DET2035 grown on CYS80 were subjected to high resolution MS/MS, using CID at 35eV. MSFinder 3.44 was then utilized to confirm the molecular formulas of the annotated ions, based on MS/MS fragmentation data.

Inspection of the UPLC-HRMS fraction data revealed that the fraction(s) containing **Unknown 1**, also possessed additional impurities. As such, the fractions were subjected to another round of prep-HPLC fractionation. The protocol utilized to fractionate fraction 14 of DET2035 further (DET2035_F14) was a 16-minute run where the gradient began at 70% ACN and the flow rate was set to 20mL/min. By the 1.0-minute time point, the gradient was increased to 80% ACN and then increased again by the 2-minute mark to 85% ACN. The gradient was steady at 85% ACN until the 6-minute mark, at which point, the gradient was to be 90% ACN. The gradient then increased to 95% ACN by the 11-minute mark, before being decreased to a 5% gradient by the 13-minute mark. The gradient was held at 5% for 1 minute, and then given 2 minutes to re-equilibrate to 70% ACN. The fractions resulting from the second prep-HPLC run were then prepared for UPLC-HRMS analysis in the same manner as prior (to 500 μg/mL).

Inspection of the UPLC-HRMS data, once again, suggested there were impurities in the fraction. As such, the fraction containing **Unknown 1** was subjected to a final prep-HPLC run. The protocol a gradient of 20% ACN at time-point 0, and a linear increase to 85% ACN by the 15-minute mark (at a flow rate of 30 mL/min). The fractions were submitted for a final UPLC-

HRMS analysis, which showed **Unknown 1** and a single impurity. To obtain a pure sample of **Unknown 1**, an orthogonal separation was used.

3.2.10 Final Compound Purification

Analytical thin layer chromatography (TLC) was carried out on TLC plates (about 5 x 5 cm with 0.25 mm thickness, silica gel GF_{254}, Merck) from the commercially available sheets. An aliquot of crude extract was spotted onto the silica gel plate and allowed to dry for about a minute. After the spot had dried onto the plate, the plate was developed in several solvent systems as the mobile phase, based on systems reported in literature (Caputo and Viola, 1977; Robeson and Strobel, 1981; Jiang et al., 2008; Xie et al., 2009). The plates were developed in a glass chamber (at room temperature), with the solvent systems until the solvent was observed to elute to about 1 cm from the top of the plate. The developed plates were dried under normal air and the spots visualized under UV light. The spots were circled and the solvent front was noted, so as to allow for the calculation of the retention factor (R_f) of the spots. The mobile phase of 5% MeOH in chloroform was selected for compound purification since it provided the best separation.

In order to obtain a pure sample of the compound of interest (**Unknown 1**), preparative TLC was carried out again, though on a larger TLC plate (20 x 20 cm with 0.25 mm thickness, silica gel GF_{254}, Merck). About 20 mg of the crude extract was loaded onto the plate by slowly dispensing it along the TLC plate (about 2.5 cm from the bottom of the plate) and allowed to dry for a few minutes. The plate was then developed with 5% MeOH in chloroform as the mobile phase for a period of about 20 minutes. UV illumination of the plate showed further separation was needed. The plate was dried under normal air and placed back into the TLC chamber and allowed to run

for another 40 minutes. The developed plate was dried and visualized under UV light. The R_f values of the isolated bands were calculated and compared.

The bands of the TLC plate were individually scraped off with a razor blade. The samples were dissolved in about 50 mL of EtOAc and placed on a magnetic stirrer until the silica had completely broken up in solution. After this point, a column was loaded with about 5cm of acid-washed sand, rinsed with EtOAc, and the scraped sample solution poured into the column. Air was passed through the column to keep the solution mobile and the samples were collected in round-bottomed flasks. After all solution had been collected, they were dried down on a rotary evaporator, and transferred into pre-weighed scintillation vials.

3.2.11 NMR

^{1}H-NMR, ^{13}C-NMR, and HSQC data were obtained using a Bruker Avance HD 600 of the NMR Facility of the University of Ottawa by Mike Darnowski of the Boddy Laboratory. The ^{1}H-NMR data was obtained at 600MHz and ^{13}C-NMR data was obtained at 150MHz. The sample submitted for NMR analyses was **Unknown 1,** which was believed to be dehydrocurvularin (DHC) on the basis of having a *m/z* matching what is reported in literature. So as to allow for direct comparison to literature, **Unknown 1** was re-suspended in CD_3OD for NMR analyses.

3.3 Results

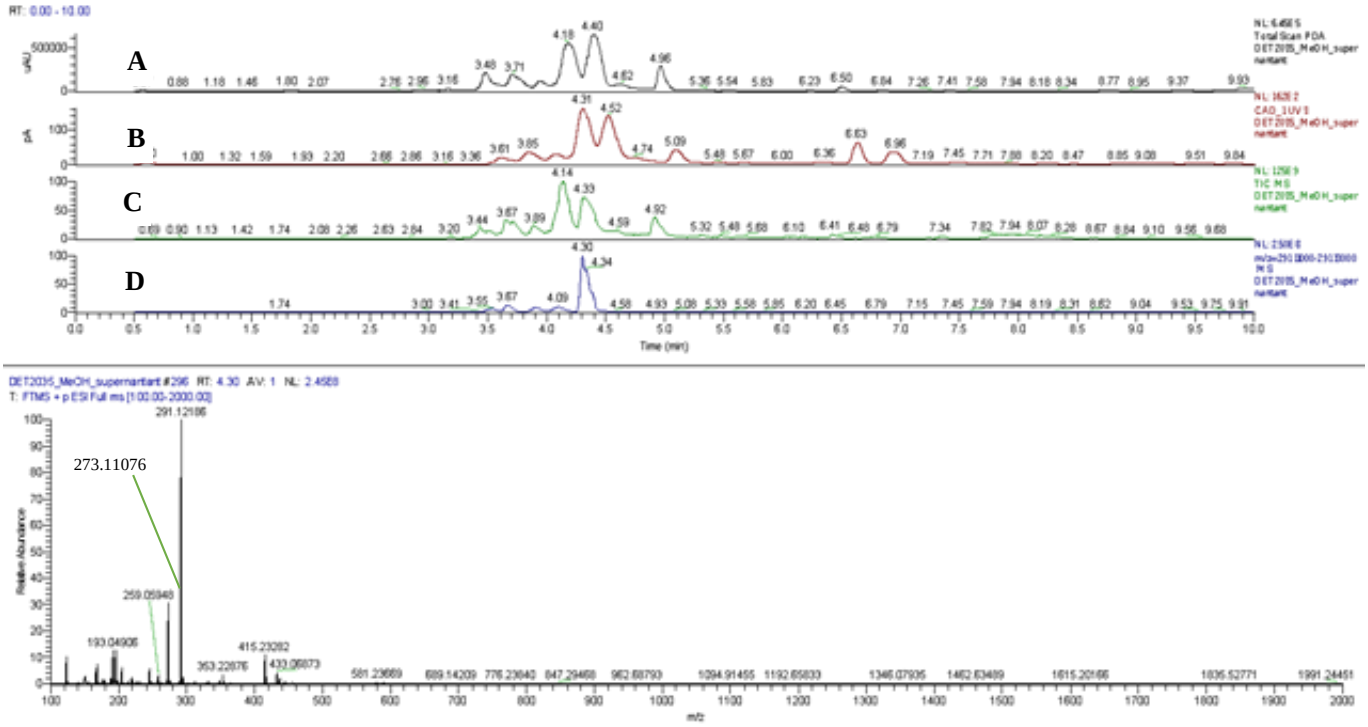

Figure 3.3. The spectra obtained from a UPLC-HRMS run, in which, the DET2035 bulk extract was analyzed. The photo diode array (PDA) (**A**), charged aerosol detector (CAD) (**B**), total ion chromatograph (TIC) (**C**), extracted ion chromatogram (XIC) (**D**) and MS spectrum (**E**) were monitored for each run and exemplified herein.

The scale-up fermentation growth was conducted by inoculating DET20235 on 300 plates of CYS80 (the media type which was observed to promote the maximal production of the **Unknown 1**). Following the usual 2-week incubation period, the plates were harvested, and a bulk extract was generated. A portion of the bulk extract was prepared for UPLC-HRMS so as to ensure the family of **Unknowns** were present in sufficient relative quantity (**Figure 3.3**). As exemplified in **Figure 3.3**, the features expected at RT3.7, RT3.91, RT4.14, RT4.33 (RT4.38 in previous chapter), and RT4.47 were present in the sample (in the total ion chromatogram (TIC)), and validated by the CAD spectrum (collisionally activated dissociation) (**Figure 3.3B**).

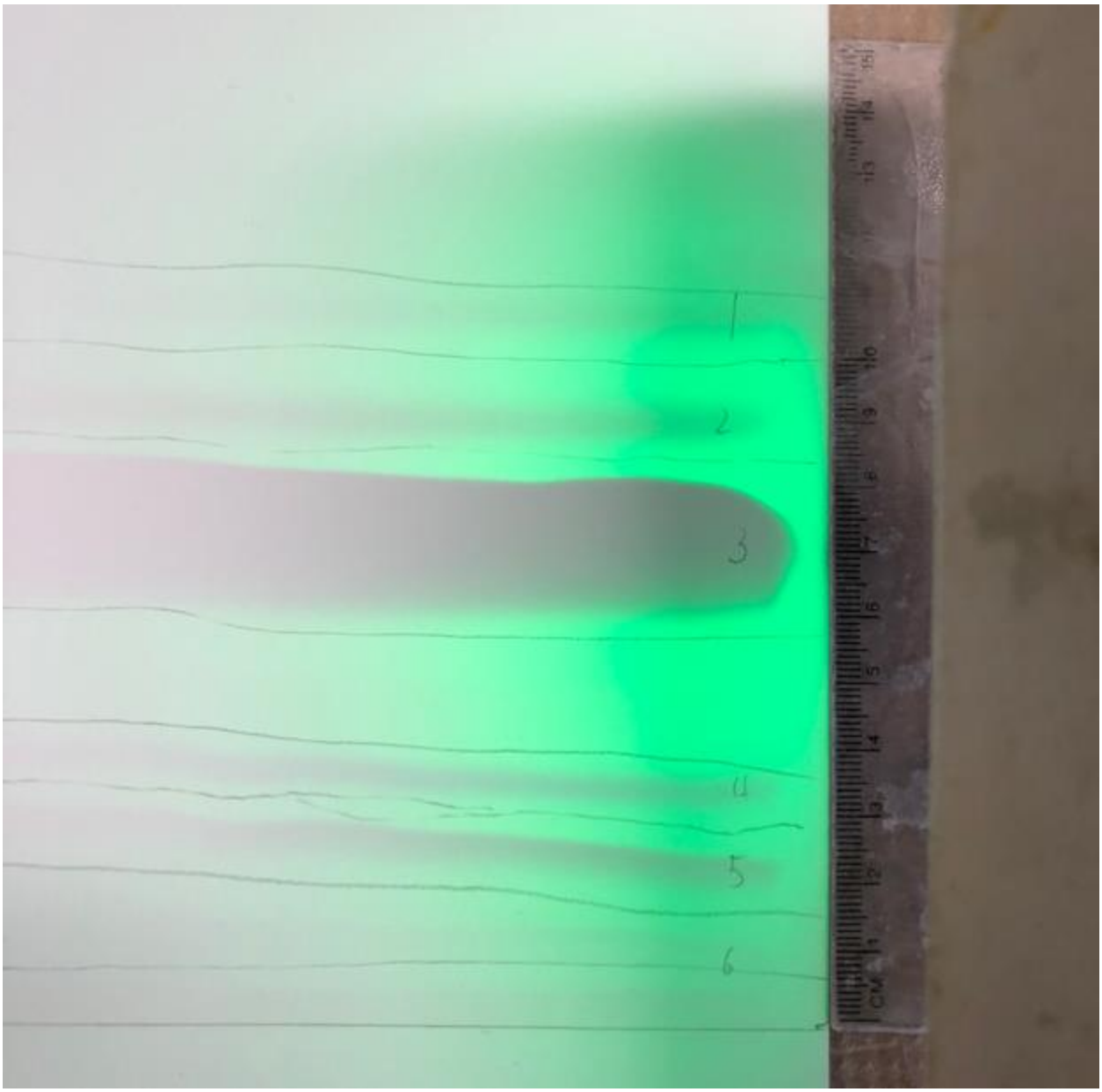

Figure 3.4. The analytical thin layer chromatography (TLC) plate (20x20cm with 0.25mm thickness, silica gel), loaded with ~20mg of a the fraction containing Unknown 1, and bands visualized using visible light.

In order to obtain a pure sample of **Unknown 1** from the bulk extract, an orthogonal purification strategy was utilized. The bulk sample was first fractionated using a prep-HPLC, and then subjected to preparative TLC in a mobile solvent of 5% MeOH in Chloroform. The preparative TLC on a semi-pure fraction containing **Unknown 1** is shown in **Figure 3.4**. **Unknown 1** is

labelled band 3. The other bands on the plate are impurities present in the sample, which may be structurally related to **Unknown 1** as they share similar chemical properties (evidenced by their comparable chemical separations). The retention factor (R_f) obtained for each of the bands is as follows: band **6** had a R_f of 0.07, **5** had a R_f of 0.13, **4** had a R_f of 0.22, **3** had a R_f of 0.47, **5** had a R_f of 0.6, and **6** had a R_f of 0.7. There was sufficient resolution to obtain a sample of the compound that each of the bands denote, though band **3** was the only sample where sufficient material (8.7 mg) was recovered for NMR-analyses.

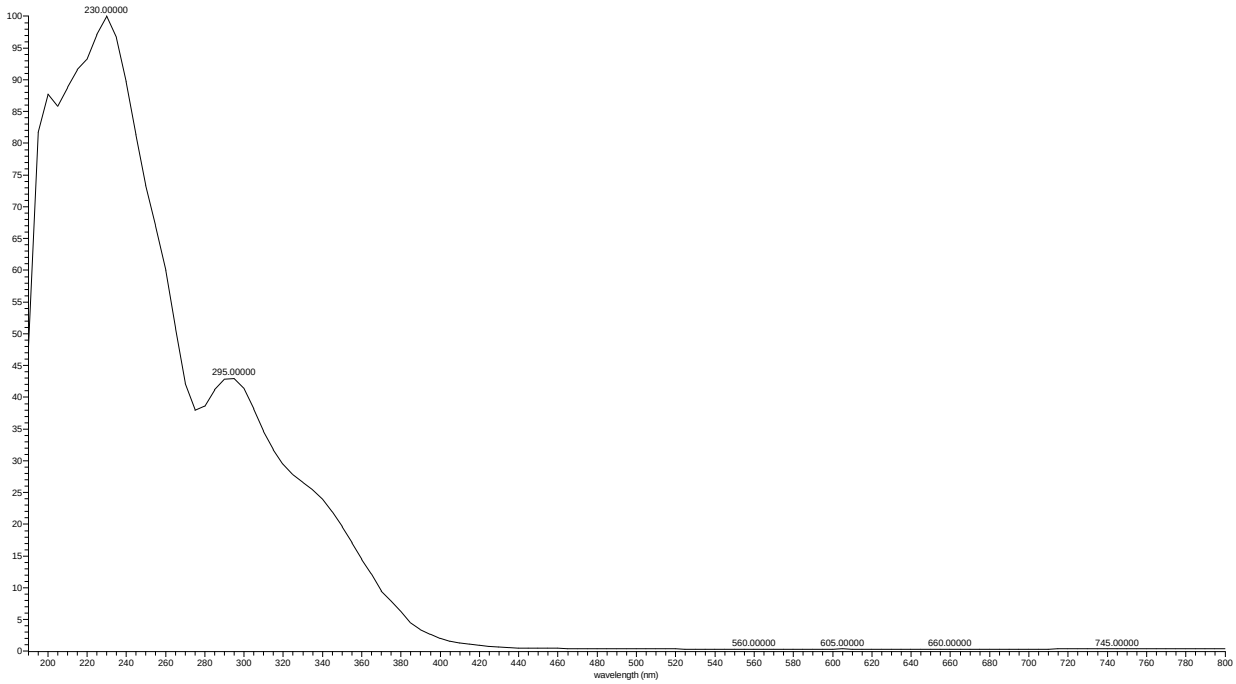

Figure 3.5. The UV spectrum (utilizing a photo diode array (PDA) detector, monitoring the wavelengths of 190-800nm) for the bulk extract, at the time where Unknown 1 eluted from the column (and hit the detector).

The UV absorbance spectrum of the **Unknown 1** ion, measured in a gradient of ACN and H_2O, is shown in **Figure 3.5**. The UV absorbance spectrum reveals UV absorbance peaks at 205, 230, 295nm, and a hump at about 335 nm. These values are consistent with DHC reported in literature (Jiang et al., 2008).

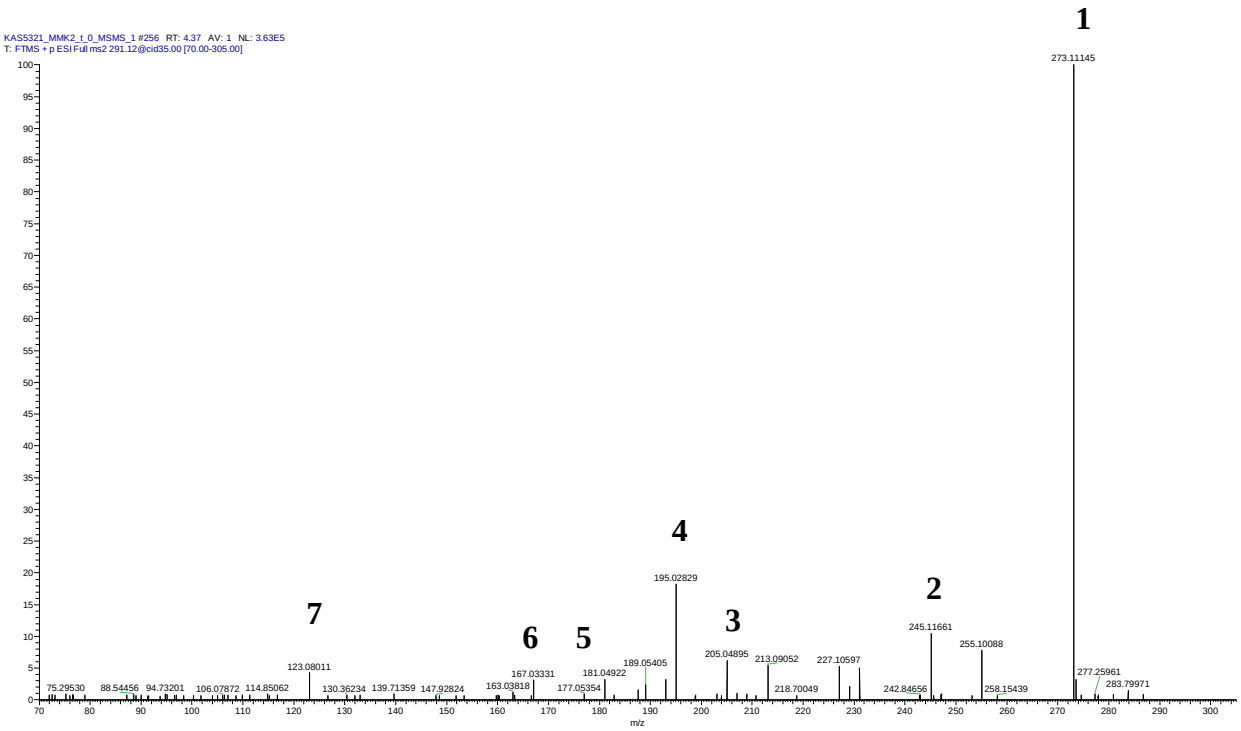

Figure 3.6. The MS/MS spectrum resulting from the fragmentation of Unknown 1, at a CID of 35eV.

The MS/MS spectrum resulting from the fragmentation of **Unknown 1** is shown in **Figure 3.6**. The MS/MS analysis revealed the presence of several peaks that support the annotation of the ion as DHC. Namely, the presence of the peaks with an associated *m/z* of 273.11145 (**Figure 3.6.1**), 245.11661 (**Figure 3.6.2**), and 205.04895 (**Figure 3.6.3**), 195.02829 (**Figure 3.6.4**), 177.05354 (**Figure 3.6.5**), 167.03331 (**Figure 3.6.6**) and 123.08011 (**Figure 3.6.7**). These peaks support the annotation of 291.12239 as DHC. **1** is consistent with the neutral loss of H_2O, **2** with a loss of a CO from **1**, and **3** with the loss of C_3H_4 from **2**. The presence of the fragments with a *m/z* of 273 (**1**), 195 (**4**), 167 (**6**), and 123 (**7**) have been reported for DHC in literature (Hassan, 2007; Jiang et al., 2008).

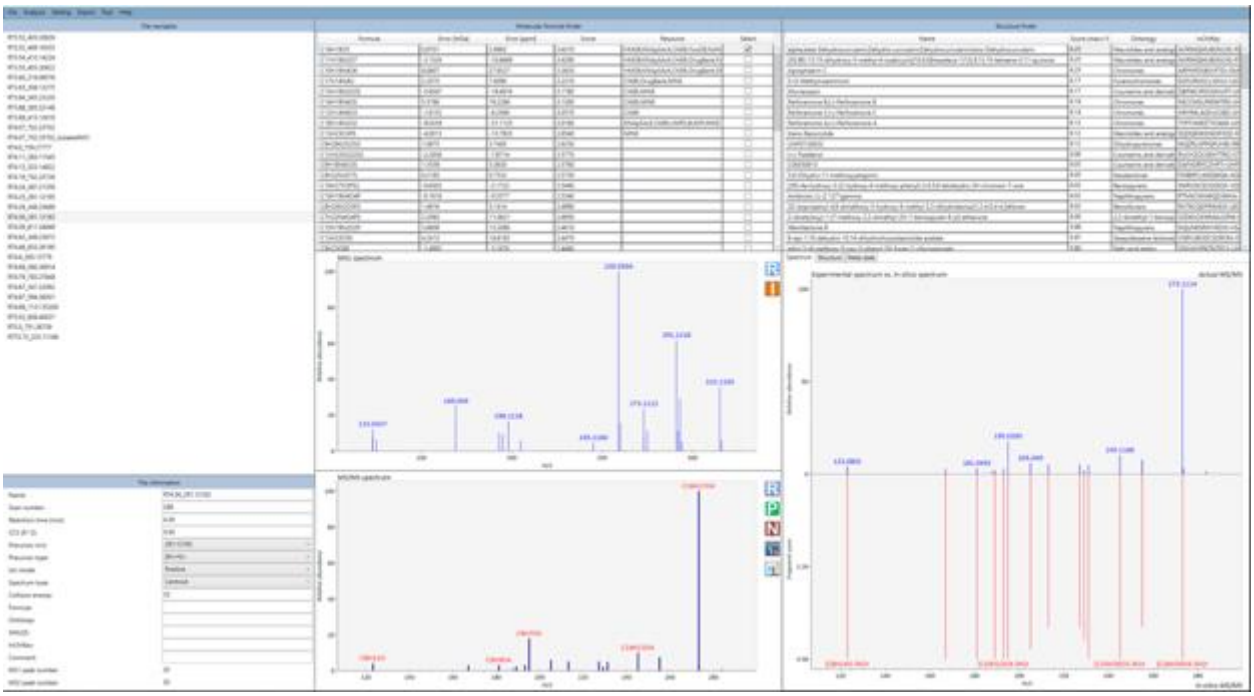

Figure 3.7. Utilizing the MS1 and MS/MS data in MSFinder (ver. 3.44) to confirm the annotation of Unknown 1 as DHC.

The MS1 and MS/MS spectra were utilized in MSFinder to assess the validity of the annotation of RT4.33_291.12239 as DHC (**Figure 3.7**). MSFinder compares the fragmentation pattern of compounds inputted by the user to that of publicly available databases. Through use of the isotope pattern observed in the MS1 spectrum, and the MS/MS spectral data, the results of the MSFinder analysis serve to support the annotation of RT4.33_291.12239 as DHC. Specifically, DHC was revealed to be 2.992 ppm from the mass obtained (291.12239) and each fragment observed in the MS/MS spectrum inputted into the software shows a high fragment score to that of DHC.

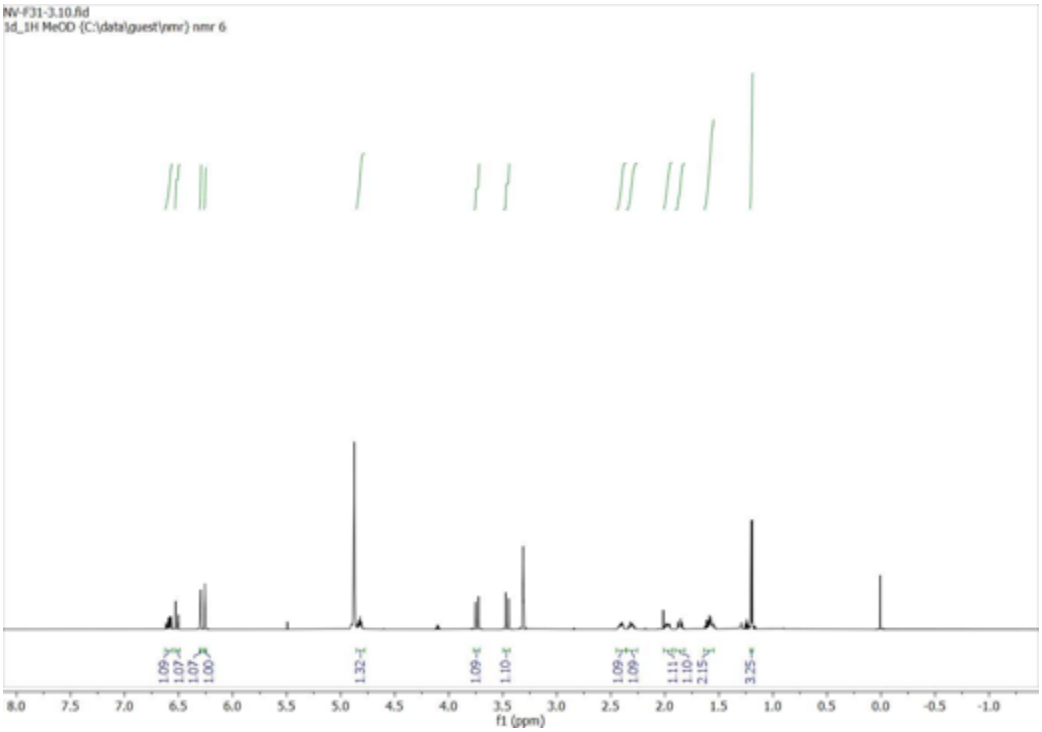

Figure 3.8. The ^{1}H-NMR of Unknown 1 in DHC (obtained using a Bruker Avance HD 600, at 600MHz).

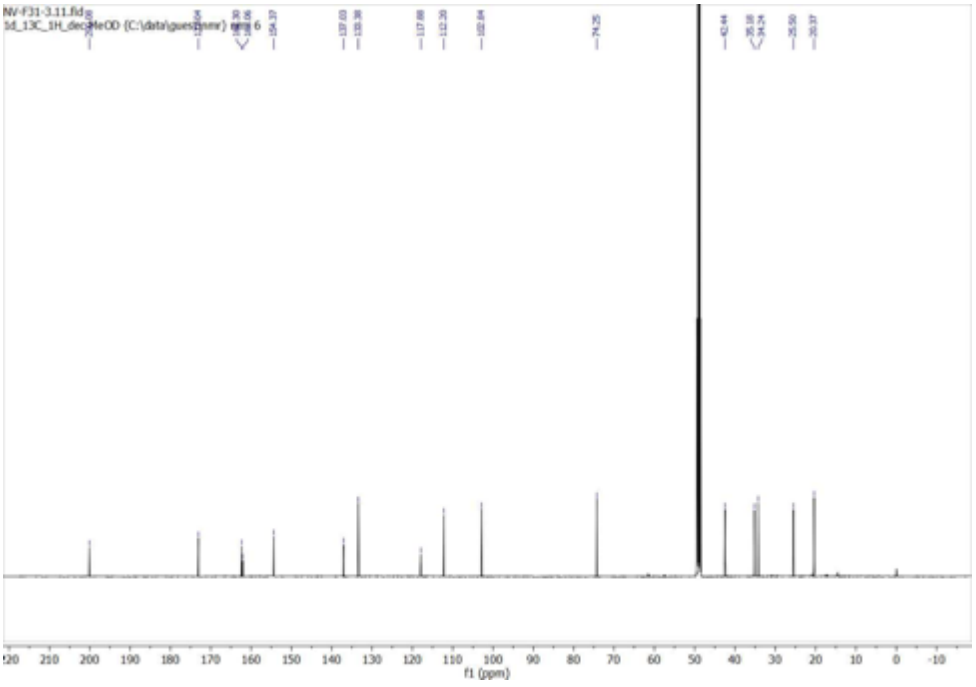

Figure 3.9. The ^{13}C-NMR of DHC in MeOD (obtained using a Bruker Avance HD 600, at 150MHz).

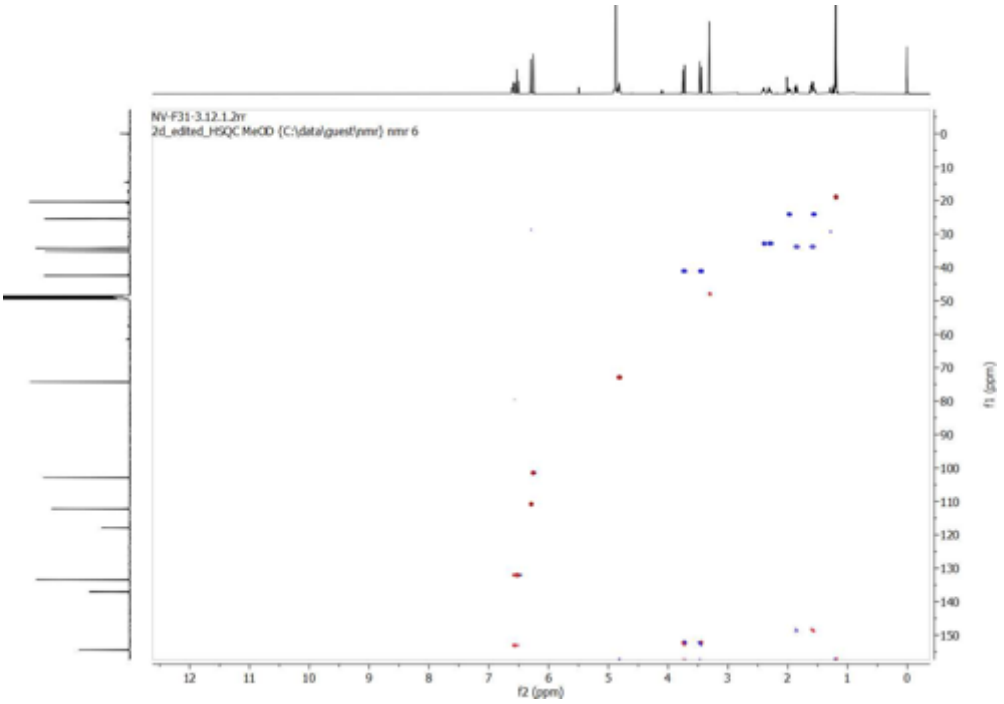

Figure 3.10. The HSQC of DHC in MeOD (obtained using a Bruker Avance HD 600), with the ^{1}H-NMR being displayed on the x-axis and the ^{13}C-NMR being displayed on the y-axis.

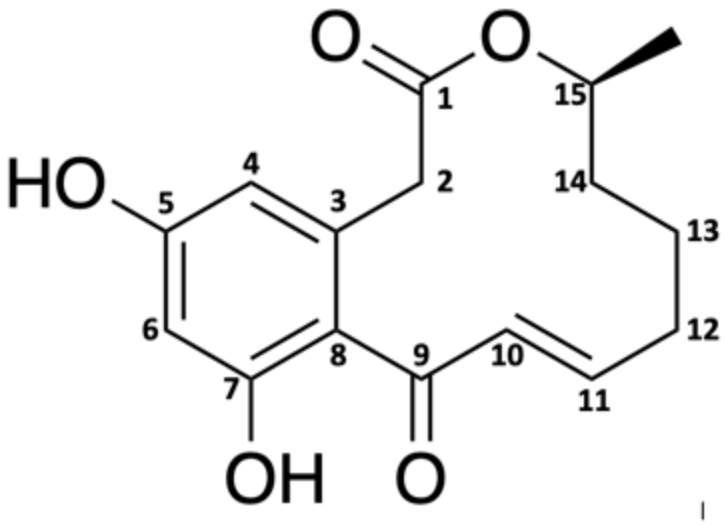

Figure 3.11. The structure of dehydrocurvularin (DHC), where each carbon has been assigned a number utilized in annotation

Table 3.1. A summary of the ^{1}H-NMR data obtained for Unknown 1 in MeOD (obtained using a Bruker Avance HD 600, at 600MHz)

H number*	δ_H (ppm) (multiplicity, coupling (Hz), integral)	δ_H (ppm) (multiplicity, coupling (Hz), integral) (Hassan, 2007)
1		
2A	3.74 (d, J = 16.6Hz, 1H)	3.72 (d, J = 16.6Hz, 1H)
2B	3.46 (d, J = 16.6Hz, 1H)	3.44 (d, J = 16.6Hz, 1H)
3		
4	6.30 (d, J = 2.3Hz, 1H)	6.28 (d, J = 2.0Hz, 1H)
5		
6	6.26 (d, J = 2.3Hz, 1H)	6.24 (d, J = 2.0Hz, 1H)
7		
8		
9		
10	6.51 (dt, J = 15.6, 1.1Hz, 1H)	6.49 (d, J = 15.4Hz, 1H)
11	6.59 (ddd, J = 15.5, 8.4, 5.7Hz, 1H)	6.57 (ddd, J = 15.4, 8.2, 5.6Hz, 1H)
12A	2.41 (ddddd, *J* = 14.7, 7.4, 5.8, 2.7, 1.4 Hz, 1H)	2.38 (m, 1H)
12B	2.36-2.26 (m, 1H)	2.29 (m, 1H)
13A	2.01-1.93 (m, 1H)	1.96 (m, 1H)
13B	1.64-1.55 (m, 2H)**	1.56 (m, 1H)
14A	1.91-1.81 (m, 1H)	1.85 (m, 1H)
14B	1.64-1.55 (m, 2H)**	1.58 (m, 1H)
15	4.82 (tdd, J = 12.8, 6.2, 4.2Hz, 1H)	4.82 (m, 1H)
CH_3	1.20 (d, J=6.4Hz, 3H)	1.18 (d, J=6.6Hz, 3H)

* = Numbering of atoms is in accordance with literature (Hassan, 2007) and is exemplified in **Figure 3.11**.

** = two peaks are indistinguishable from one another so are integrated together. 1H belongs to 13B (closed to 1.55) and the other 1H belongs to 14B (closer to 1.64).

Table 3.2. A summary of the ^{13}C-NMR data obtained for DHC in MeOD (obtained using a Bruker Avance HD 600, at 150MHz)

C number*	δc (ppm)	δc (ppm) (Hassan, 2007)
1	173.0	173.0
2	42.4	42.4
3	137.0	137.0
4	112.2	112.2
5	162.0	162.0
6	102.8	102.8
7	162.3	162.2
8	117.9	117.8
9	200.1	200.0
10	133.4	133.3
11	154.4	154.3
12	34.2	34.2
13	25.5	25.4
14	35.2	35.1
15	74.3	74.2
CH_3	20.4	20.3

* = Numbering of atoms is in accordance with literature (Hassan, 2007) and is exemplified in **Figure 3.11**.

To assess the purity of the sample of **Unknown 1** following its orthogonal purification, and whether it was, indeed, DHC, it was submitted for NMR analyses. The ^{1}H-NMR, ^{13}C-NMR, and HSQC results are shown in **Figure 3.8, Figure 3.9,** and **Figure 3.10**, respectively. A tabular summary of the ^{1}H-NMR and ^{13}C-NMR are shown in **Table 3.1** and **Table 3.2,** respectively.

The ^{1}H-NMR showed **Unknown 1** to be of sufficient purity for characterization. Analysis of the spectra revealed two aromatic protons (δ_H of 6.30 ppm and δ_H of 6.30 ppm) as well as two alkene protons (δ_H of 6.51 ppm and δ_H of 6.59 ppm). The alkene protons show a large coupling constant (J=15.4 Hz), suggesting a *trans* olefin. Two olefinic protons form the AB portion. Protons at δ_H 3.74 ppm and 3.45 ppm each have identical coupling constants (J=16.6 Hz),

consistent with geminal coupling. The three protons at δ_H of 1.20 ppm are a doublet with J=6.4 Hz, consistent with a methyl group coupled to a single adjacent hydrogen.

Inspection of the ^{13}C-NMR revealed 6 aromatic carbons (C3-C8 in **Table 3.2**). δ_C of 162.02 and 162.30 ppm are consistent with aromatic carbons experiencing the added de-shielding of being bound to an OH group. The resonance of 173.04 ppm (C1) is consistent with a carbonyl carbon that is experiencing de-shielding from at least one neighbouring heteroatom (an ester group). Furthermore, the resonance at 72.4 (C15) is consistent with a saturated carbon bound to the oxygen. Combined, these two signals suggest an ester. The observed resonance at 200.08 ppm (C9) suggests there is a ketone in the structure. The resonances observed at 133.4 ppm and 154.4 ppm (C10 and C11, respectively) are consistent with an olefin connected to an electron withdrawing group. With the ketone, this suggests an α,β - unsaturated ketone is present. The remaining resonances of 34.2 (C12), 25.5 (C13), 35.2 (C14), and 20.4 (CH3) are all consistent with alkane chain carbons.

Analysis of the HSQC confirms that the carbons at δ_C of 200 ppm, 173 ppm, 137.3 ppm, 162.02, 162.30, and 117.88 ppm have no attached hydrogens. The carbons are 112.20 ppm and 102.8 ppm are attached to the aromatic protons at δ_H of 6.30 and 6.26 respectively. The alkene protons at δ_H of 6.50 ppm and 6.59 ppm are connected to the carbons at 133.4 and 154.4, respectively. Lastly, the carbon δ_C 74.3 ppm is indicative of a carbon with a single bond to an electronegative heteroatom and is connected to the proton at δ_H 4.82 ppm.

Four alkane carbons are clearly identifiable. The carbon at δ_C 42.4 ppm couples to the protons at δ_H 3.74 ppm and 3.46 ppm. The carbon at δ_C 35.2 ppm couples to the protons at δ_H 1.91 – 1.81 ppm and 1.64 – 1.55 ppm. The carbon at δ_C 34.2 ppm couples to the protons at δ_H 2.41 ppm and 2.36 – 2.26 ppm. Finally, the methyl carbon at δ_C 20.4 ppm couples to three protons at δ_H 1.20 ppm.

These data are consistent with the functional groups found in DHC, a tetrasubstituted electron, a rich aromatic, an α,β - unsaturated ketone, an ester, four aliphatic groups and a single methyl group. The assignment is further supported by comparison of NMR data to literature NMR data as seen in **Table 3.1** and **Table 3.2** (Hassan, 2007; Aly et al., 2010).

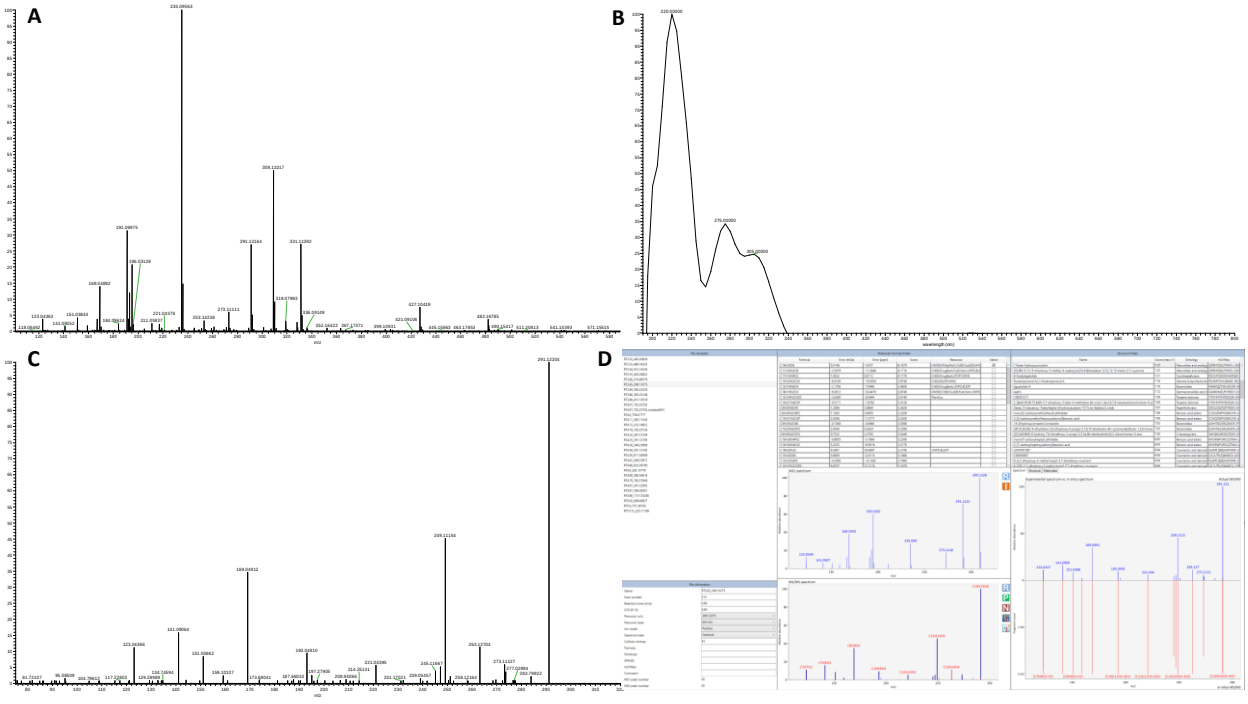

Figure 3.12. The (A) MS1, (B) UV-spectrum, and (C) MS/MS, along with (D) the MSFinder results for Unknown 2, the ion suspected to be 11-(α or β)-hydroxycurvularin.

Figure 3.12 represents the MS1 spectrum (**A**), UV spectrum (**B**), MS/MS spectrum (**C**) and MSFinder results (**D**) obtained for **Unknown 2**. **Figure 3.12A** revealed that the parent pseudomolecular ion was 309.13217 (based on the presence of the water loss and sodiated adduct, represented by 291.12164 and 331.11392, respectively). Comparison of the *m/z* of 309.13217 to an internally compiled database revealed that **Unknown 2** could, therefore, be annotated as 11-hydroxycurvularin. 11-hydroxycurvularin is a compound that can have the hydroxy group at the 11-position in either the alpha- (α) or beta- (β) conformation. Resolution between which is present in the bulk sample is not possible as the two have the same mass. The UV absorbance spectrum of the ion annotated as 11-(α or β)-hydroxycurvularin, measured in a gradient of ACN and H_2O, is shown in **Figure 3.12B**. The UV absorbance spectrum revealed a UV absorbance peaks at 220, 275, and 305nm, consistent with what has been reported in literature (Lai et al., 1989). The MS/MS fragmentation pattern of the ion annotated as 11-(α or β)-hydroxycurvularin is shown in **Figure 3.12C**. Key fragments of the spectrum have a m/z of 291, 273, 245, 193, 169, and 123, which are consistent with the DHC fragments observed in **Figure 3.6**. Finally, the results of inputting the MS1 and MS/MS data into MSFinder reveals that by comparison of the *m/z* to that expected for $C_{16}H_{20}O_6$, the isotope pattern, and the fragmentation to publicly available databases, the feature can be annotated as 11β-hydroxycurvularin (**Figure 3.12D**).

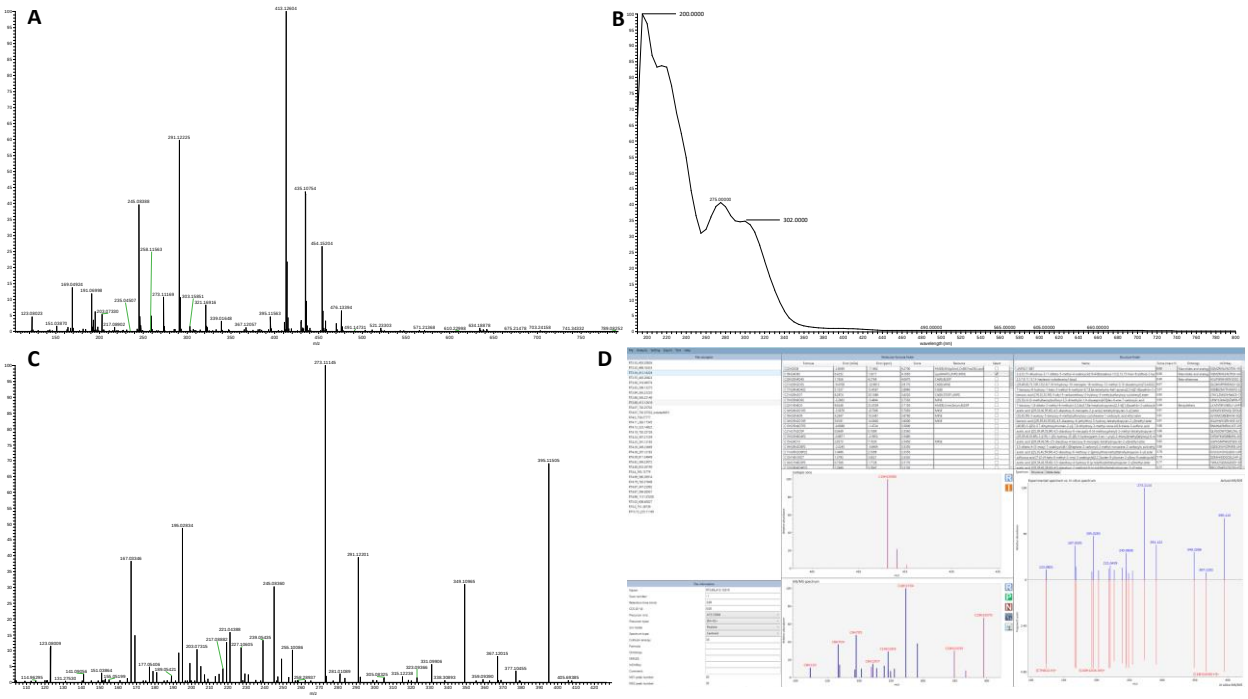

Figure 3.13. The (A) MS1, (B) UV-spectrum, and (C) MS/MS, along with (D) the MSFinder results for Unknown 3, the ion suspected to be Sumalarin C.

Figure 3.13 represents the MS1 spectrum (**A**), UV spectrum (**B**), MS/MS spectrum (**C**) and MSFinder results (**D**) obtained for **Unknown 3. Figure 3.13A** revealed that the parent pseudomolecular ion was 413.12604 (based on the presence of the water loss and sodiated adduct, represented by *m/z* 395.11563 and *m/z* 435.10754, respectively). Comparison of the *m/z* of 413.12604 to an internally compiled database revealed that the **Unknown 3** could, therefore, be annotated as Sumalarin C. The UV absorbance spectrum of the ion annotated as Sumalarin C, measured in a gradient of ACN and H_2O, is shown in **Figure 3.13B**. The UV absorbance spectrum revealed a UV absorbance peaks at 200, 220, 275, and 302nm, consistent with the UV spectrum values reported in literature (excluding 220nm) (Meng et al., 2013). The MS/MS fragmentation pattern of the ion annotated as Sumalarin C is presented in **Figure 3.13C**. Key fragments of the spectrum were *m/z* 395, 291, 273, 245, 203, 195, 167, and 123. Inputting the obtained MS1 and MS/MS data into MSFinder showed a strongest match to a molecular formula

of $C_{18}H_{24}O_8S$ (with a mass difference of 1.0317 ppm), in which Sumalarin C was the strongest match (denoted by both the UNPD ID and IUPAC name in the top right of the window) (**Figure 3.13D**).

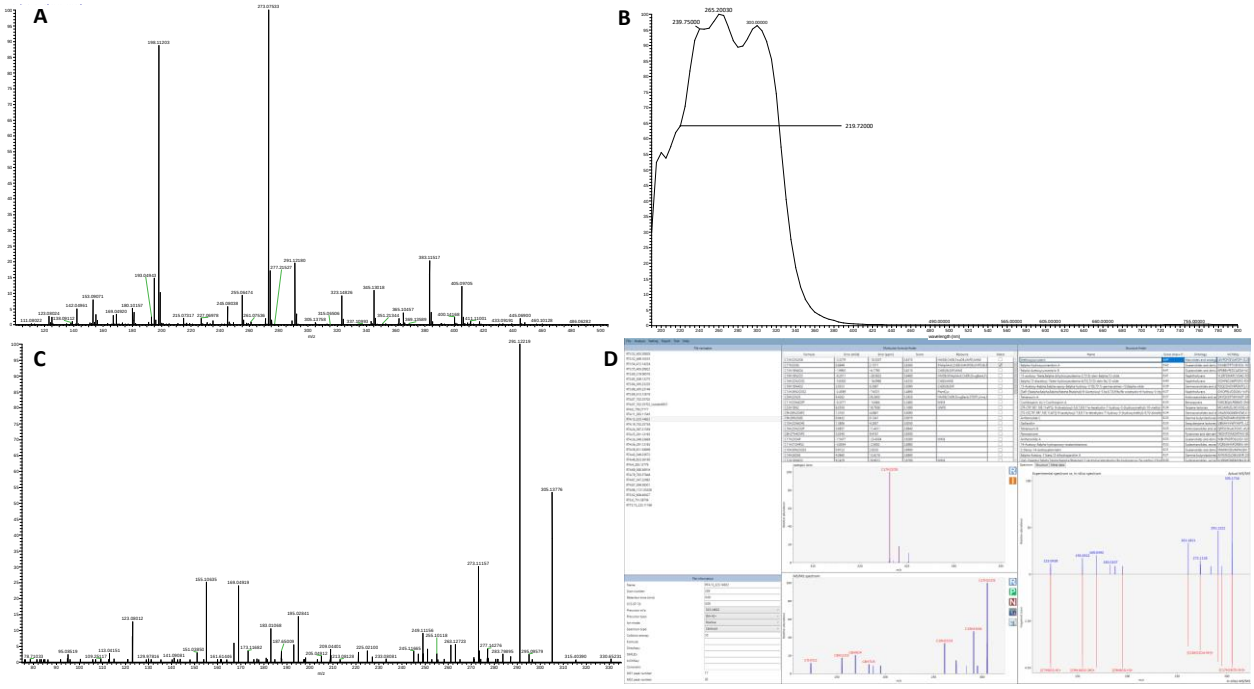

Figure 3.14. The (A) MS1, (B) UV-spectrum, and (C) MS/MS, along with (D) the MSFinder results for Unknown 4, the ion suspected to be 11-(α or β)-methoxycurvularin.

Figure 3.14 represents the MS1 spectrum (**A**), UV spectrum (**B**), MS/MS spectrum (**C**) and MSFinder results (**D**) obtained for **Unknown 4. Figure 3.14A** revealed that the parent pseudomolecular ion was 323.14826 (based on the presence of the water loss and sodiated adduct, represented by 305.13758 and 345.13018, respectively). Comparison of the *m/z* of 323.14826 to an internally compiled database allowed **Unknown 4** to be annotated as 11-(α or β)-methoxycurvularin. The UV absorbance spectrum of the ion annotated as 11-(α or β)-methoxycurvularin, measured in a gradient of ACN and H_2O, is shown in **Figure 3.14B**. The UV

absorbance spectrum revealed a UV absorbance peaks at 200, 220, 240, 265, and 300nm. This is generally in accordance with what is reported in literature, differing by, at most, 10nm (Kobayashi et al., 1988). The MS/MS fragmentation pattern of the ion annotated as 11-(α or β)-methoxycurvularin may be denoted by **Figure 3.14C**. Key fragments of the spectrum are *m/z* 305, 291, 273, 245, 205, 195, 169, 167, and 123, which are consistent with the fragments observed in the curvularins/other unknowns. In this way, the MS/MS results serve to confirm that the compound in the sample is of the curvlarins family. Finally, the results of inputting the MS1 and MS/MS data into MSFinder reveals that by comparison of the obtained *m/z* to that expected with a molecular formula of $C_{17}H_{22}O_6$, the isotope pattern, and the fragmentation to publicly available databases, the feature can be annotated as methoxycurvularin (**Figure 3.14D**).

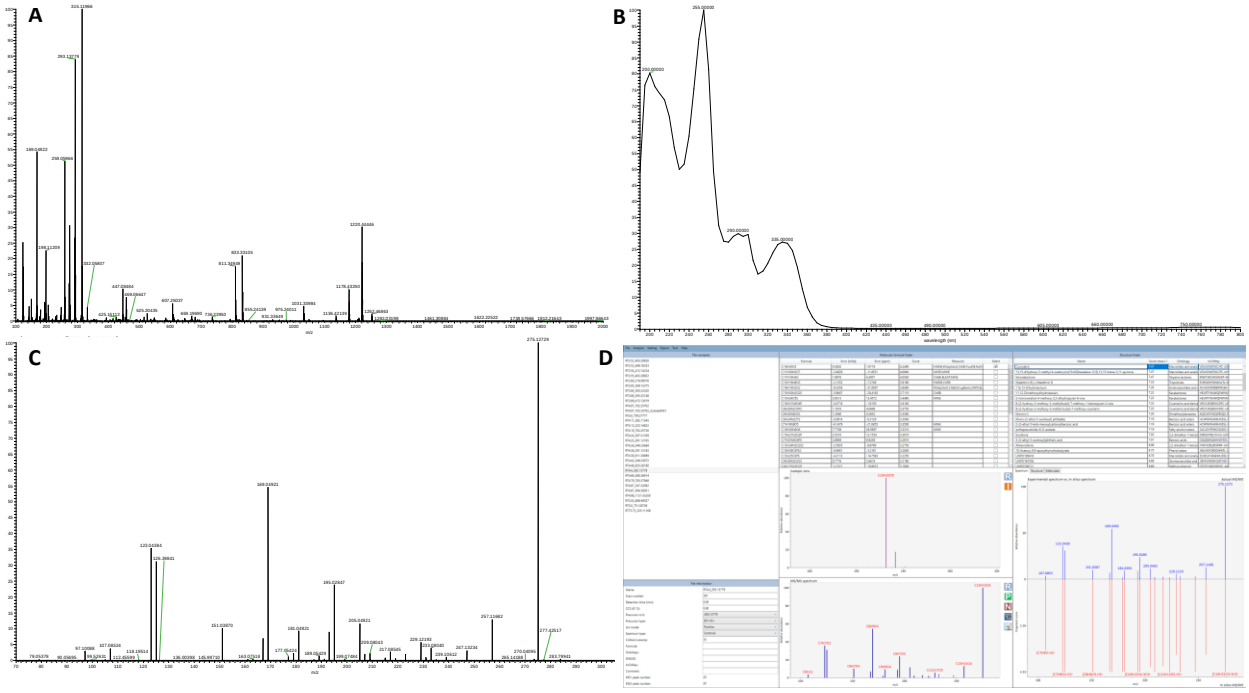

Figure 3.15. The (A) MS1, (B) UV-spectrum, and (C) MS/MS, along with (D) the MSFinder results for Unknown 5, the ion suspected to be Curvularin.

Figure 3.15 represents the MS1 spectrum (**A**), UV spectrum (**B**), MS/MS spectrum (**C**) and MSFinder results (**D**) obtained for **Unknown 5. Figure 3.15A** revealed that the parent pseudomolecular ion was 293.13779 (based on the presence of the water loss and sodiated adduct, represented by 275.12729 and 315.11956, respectively). Comparison of the *m/z* of 293.13779 to an internally compiled database allowed **Unknown 5** to be annotated as curvularin. The UV absorbance spectrum of the ion annotated as curvularin, was measured in a gradient of ACN and H_2O (**Figure 3.15B**). The UV absorbance spectrum contained UV absorbance peaks at 200, 255, 290, 300, and 335nm. UV absorbance peaks in these ranges are not consistent with literature (Kobayashi et al., 1988). The MS/MS fragmentation pattern of the ion annotated as curvularin is shown in **Figure 3.15C**. Key fragments of the spectrum are 275, 247, 205, 195, 177, 169, 167, 125, and 123, which are consistent with the fragments observed in the curvularins discussed previously. In this way, the MS/MS results serve to instantiate that the compound in the sample is of the curvularins family. Finally, the results of inputting the MS1 and MS/MS data into MSFinder reveals that by comparison of the obtained *m/z* to that expected with a molecular formula of $C_{16}H_{20}O_5$, the isotope pattern, and the fragmentation to publicly available databases, the feature can be annotated as curvularin (**Figure 3.15D**).

Table 3.3. The suspected metabolite identity, suspected molecular formula, chemical classification and mass error (ppm) of several key mass features produced by DET2035 and detected in the bulk-extract.

Mass Feature	Suspected Metabolite	Molecular Formula	Mass Error (ppm)	Chemical Classification
RT3.70_309.13205	11-(α or β)-hydroxycurvularin	$C_{16}H_{20}O_6$	1.6707	Macrolides and Analogues (Lactone)
RT3.9_413.12619	sumalarin C	$C_{19}H_{24}O_8S$	1.0317	Macrolides and Analogues (Lactone)
RT4.15_323.14822	11-(α or β)-methoxycurvularin	$C_{17}H_{22}O_6$	2.1571	Macrolides and Analogues (Lactone)
RT4.35_291.12239	dehydrocurvularin	$C_{16}H_{18}O_5$	2.9992	Macrolides and Analogues (Lactone)
RT4.40_293.13770	curvularin	$C_{16}H_{20}O_5$	1.9176	Macrolides and Analogues (Lactone)

A summary of the annotation of the key mass features utilized in the multi-variate and uni-variate statistical analyses performed in **Chapter 2** of the present thesis, is found in **Table 3.3.** The annotation of the mass features was informed by the MS, MS/MS, and UV spectra presented previously, along with the use of the MSFinder software. The annotation of the mass features as their respective metabolite was also consistent with the exact mass of the molecular ion which was less than 3 ppm different from the expected mass. Lastly, DHC was unambiguously confirmed by isolation and ^{1}H and ^{13}C NMR analysis.

3.4 Discussion

The purpose of the experimentation carried out for the present chapter was to perform a scale-up growth fermentation of isolates shown to produce unique mass features in **Chapter 2**, with the intent to isolate, purify, and elucidate the identity of the metabolites corresponding to those mass features. This was accomplished through a large-scale fermentation of the DET2035 strain. After large-scale fermentation, a bulk extract was generated and the SMs that distinguished DET2035, KAS5321, and KAS5743 as unique in the untargeted metabolomic analyses were isolated, purified, characterized and their structures elucidated. It was hypothesized that the unique mass features were related to molecular families of analogs sharing structural similarities (specifically, that these unique mass features were the curvularins).

DHC was first identified from the ether-extract of a culture filtrate of a *Curvularia* species, which gave a high yield of a compound with a formula of $C_{16}H_{20}O_5$ (Musgrave, 1956). As it was characterized from a *Curvularia* species, the name "curvularin" was proposed (Musgrave, 1956). A decade later, the culture filtrate of a species of *Curvularia* was reported to contain, in addition to curvularin, a small quantity of a second compound, $C_{16}H_{18}O_5$ (Munro et al., 1967). Its molecular formula and its origin suggested a close relationship to curvularin that was established by hydrogenation, which converted it quantitatively into curvularin (Munro et al., 1967). This second member of the curvularin family was called "dehydrocurvularin" (DHC). A decade later, DHC was detected in an *Alternaria* isolate, along with a new member of the curvularins family, identified as β-hydroxycurvularin (Hyeon et al., 1976). About a decade after the discovery of β-hydroxycurvularin, β-methoxycurvularin was discovered in a *Penicillium* isolate (Kobayashi et al., 1988). Reports of different curvularin analogues have followed for the years that followed,

across *Alternaria*, *Aspergillus*, and *Curvularia* spp., with one of the most recent discoveries being the characterization of sumalarins A-D (Meng et al., 2013).

Curvularins are dihydroxyphenylacetic acid lactones (DALs) with structural similarities to the resorcylic acid lactones (RALs). To date, curvularins have been isolated from a number of phytopathogenic fungal species, namely *Alternaria*, *Aspergillus*, *Cochliobolus*, *Curvularia*, *Nectria*, and *Penicillium* spp. (Musgrave, 1956; Hyeon et al., 1976; Robeson et al., 1985; Ghisalberti and Rowland, 1993; He et al., 2004; Zhan et al., 2004; Gutiérrez et al., 2005). The distribution of these instances of detection of the curvularins spans most of the world (has been reported in isolates from Chile, Japan, Europe, Egypt, North America, and Korea) (Musgrave, 1956; Munro et al., 1967; Kusano et al., 2003; Gutiérrez et al., 2005; Aly et al., 2010; Jeon et al., 2010; Bashyal et al., 2017). The *Alternaria spp*. which have been found to produce the curvularins currently includes *Alternaria cinerariae*, *Alternaria zinniae*, *Alternaria macrospora*, and *Alternaria sp*. isolate from the leaves of *Acaia mangium*, and *Alternaria tomato* (which would now typically be referred to as an *Alternaria alternata* tomato pathotype isolate) (Hyeon et al., 1976; Robeson et al., 1985; Vurro et al., 1998; Jeon et al., 2010; Cochrane et al., 2015). Additionally, 2 chlorinated DHC analogues were isolated from an *Alternaria sp*. isolated from the leaf tissue of *Astragalus lentinginosus* (Bashyal et al., 2017). The findings of the present chapter represent the first isolation of DHC from an *A. arborescens* isolate. Additionally, this represents the first report of sumalarin C in an *Alternaria* species.

In this study, the UV-spectrum obtained by monitoring the DHC sample through a PDA, showed several peaks in the near-ultraviolet range. Specifically, the functional groups of the DHC

sample showed absorption at 205, 230 and 295nm (**Figure 3.5**). The obtained UV absorption spectrum was generated in a gradient of ACN and H_2O. As such, it is not in perfect accordance to literature, though this is likely to be the result of that the spectrum reported in literature was obtained in MeOH. The general shape of the obtained UV absorption spectrum, however, was in accordance with literature and the absorption peaks only deviated from the reported values by, at most, 8 ppm. As such, when the obtained UV spectrum is compared to the UV spectrum reported in literature, the annotation of the RT4.38_291.12239 as DHC can be supported.

The curvularin compounds identified in the present chapter could be assumed to be related based on the detection of several common fragments. Namely, all of the identified compounds were observed to have MS/MS spectra containing fragment ions with *m/z* values of 291, 273, 245, 205, 195, 177, 167, and 123. Sumalarin C and Curvularin were the only exceptions to this pattern. Sumalarin C was the only MS/MS spectrum which contained the ion of *m/z* of 203 instead of 205. Curvularin had the same pattern of fragments though the fragments ions were 2 mass units greater than those observed in DHC. This can be explained by that curvularin is 2 mass units greater than DHC.

In order to obtain a pure sample of DHC, an orthogonal purification system was utilized. Specifically, the bulk extract sample was separated into increasingly more pure fractions through several rounds of injection into a prep-HPLC and the resulting fraction that was indicated by UPLC-HRMS to be the least impure, was subjected to analytical TLC. Many solvent systems were tested on the 133mg fraction, according to solvent systems previously reported to be utilized to obtain pure DHC samples, ranging from 99:1 DCM:MeOH, to 5:3 EtOH:EtOAC

(Caputo and Viola, 1977; Robeson and Strobel, 1981; Jiang et al., 2008; Xie et al., 2009). Comparison of the TLC results to each other resulted in the utilization of a solvent system not previously reported in literature (5% MeOH in chloroform). Through utilization of this solvent system, an R_f in the ideal range for separation was obtained. Interestingly, the fraction separated into 6 unique bands on the TLC plate. The identity of all bands but the third is unknown. Subsequent purification work revealed that there were only trace quantities of these metabolites in the fraction (1-2mg compared to the 40mg of DHC obtained). Hydroxy-curvularin did not appear in the UPLC-HRMS trace obtained for the fraction sample (prior to it being subjected to TLC). Its trace presence on the plate could, therefore, be an artifact of DHC reacting with MeOH in solution.

As all the curvularins identified in the present chapter produced similar fragmentation spectra and a pure sample of DHC was obtained, it is likely that DHC and its analogues are a product of single gene cluster. It may be suggested that they are the product of the same BGC being activated. This is consistent with reports of the DHC BGC in literature (Xu et al., 2013). Specifically, that the DHC gene cluster is predominantly comprised of two iterative polyketide synthases, whose activity is responsible for producing DHC and all of its analogues (Xu et al., 2013). In this way, the confirmed isolation of DHC supports the identification of the other curvularins in the present work.

DHC has shown potent phytotoxicity against tomato, rice, lettuce, millet, *Zinnia elegans*, creeping thistle (*Cirsium arvense*) and weak phytotoxicity to *Xanthium occidentale* (Hyeon et al., 1976; Vurro et al., 1998; Kusano et al., 2003; Gutiérrez et al., 2005). In additional to the

phytotoxic properties DHC has exhibited on various host plant species, DHC has been reported to exhibit antifungal activity through suppression of sporulation and spore germination, antibacterial activity against several Gram-positive and Gram-negative bacteria (such as *Bacillus subtillis, Staphylococcus aureus,* and *Escherichia coli*), and significant nematicidal activity against *Meloidogyne graminicola* (Caputo and Viola, 1977; Xiang et al., 2020). DHC has also been reported to exhibit cytotoxic activity on microtubule assembly in bovine brain tubulin and garlic root tips (Kobayashi et al., 1988; Jiang et al., 2008). As such, it is currently unknown if the curvularins serve a role in virulence or if they are simply produced and upheld because their range of bioactivities confers a fitness advantage.

The DHC gene cluster has been characterized previously, in an *Aspegillus terreus* isolate (strain AH-02-30-F7), and an *A. cinerariae* (Xu et al., 2013; Cochrane et al., 2015). The architecture of the gene cluster was very similar in both, though the *A. cinerariae* contained 1 more gene than was predicted in *A. terreus* (Xu et al., 2013; Cochrane et al., 2015). Specifically, both clusters contained the two type I iterative PKSs necessary to synthesize DHC (Xu et al., 2013; Cochrane et al., 2015). In both cases, a GAL4-like fungal transcription factor is also reported between the PKSs (Xu et al., 2013; Cochrane et al., 2015). A GAL4-like transcription factor is typically a positive regulator meaning that binding to galactose induces transcription of the genes it regulates (Stone and Sadowski, 1993). Comparing this knowledge to the trends in metabolite production represented by the frequency consensus heatmap of the previous chapter (**Figure 2.14**) suggests either that the production of the curvularins by DET2035 was not media-dependent or that galactose is being derived from a source common to most media conditions utilized.

3.5 Conclusion

The purpose of the experimentation carried out for the present chapter was to build on the findings of the previous chapter by performing a scale-up growth fermentation of isolates having unique mass features, with the intent to isolate, purify, and elucidate the identity of the metabolites corresponding to those mass features. This was accomplished through a large-scale fermentation of the DET2035 strain, under the conditions which promoted the maximal production of the SMs of interest. After large-scale fermentation, a bulk extract was generated and the SMs that distinguished DET2035 as unique in the untargeted metabolomic analyses were isolated, purified, characterized and their structures elucidated. It was hypothesized that unique mass features having similar patterns in production amongst the chemical profiles of all the *Alternaria sect. A. alternata*, were related to molecular families of analogs sharing structural similarities.

The results obtained in this study serve to confirm the hypothesis and that the detected molecular "family" was the curvularins. This can be confidently concluded based on that a pure sample of DHC was obtained (proven by NMR results matching those reported in literature) (Hassan 2007; Aly et al., 2010). Additionally, the *m/z* of the mass features matched those of DHC-analogues (within 3 ppm) and the MS/MS fragments of the mass features served to suggest they are all of the same family (as per that they all share many of the same fragments). This is a significant finding as it is the first report of DHC and any DHC-analogues being isolated from an *A. arborescens* isolate. Additionally, though relying only on MS1, MS/MS and MSFinder data, this is the first report of sumalarin C in an *Alternaria* species and its second report in literature, to date. While as the approach taken in this chapter would clearly benefit the realm of natural

product discovery, its adaptation is not yet widespread. As such, this currently denotes a rare example of an untargeted metabolomics analysis being utilized to inform a targeted metabolomics analysis.

3.6 References

Alhadrami, H.A., Sayed, A.M., El-Gendy, A.O. et al. (2021). A metabolomic approach to target antimalarial metabolites in the *Artemisia annua* fungal endophytes. *Sci Rep* 11, 2770. doi:10.1038/s41598-021-82201-8

Allain, T., Groves, R., Lewis, I., and Buret, A. (2019). A163 *giardia sp*. metabolome: towards the identification of new virulence factors. *J. Can. Assoc. Gastroenterol.* 2, 323. doi:10.1093/jcag/gwz006.162.

Aly, A. H., Debab, A., Edrada-Ebel, R. A., Müller, W. E. G., Kubbutat, M. H. ., Wray, V., et al. (2010). Protein kinase inhibitors and other cytotoxic metabolites from the fungal endophyte Stemphylium botryosum isolated from Chenopodium album. *Mycosphere* 1, 153–162.

Andersen, B., Nielsen, K. F., Fernández Pinto, V., and Patriarca, A. (2015). Characterization of *Alternaria* strains from Argentinean blueberry, tomato, walnut and wheat. *Int. J. Food Microbiol.* doi:10.1016/j.ijfoodmicro.2014.11.029.

Bashyal, B. P., Wijeratne, E. M. K., Tillotson, J., Arnold, A. E., Chapman, E., and Gunatilaka, A. A. L. (2017). Chlorinated Dehydrocurvularins and Alterperylenepoxide A from *Alternaria* sp. AST0039, a Fungal Endophyte of Astragalus lentiginosus. *J. Nat. Prod.* 80, 427–433. doi:10.1021/acs.jnatprod.6b00960.

Caputo, O., and Viola, F. (1977). Isolation of a, beta-dehydrocurvularin from Aspergillus aureofulgens. *Planta Med.* 31, 31–32. doi:10.1055/S-0028-1097485.

Cochrane, R. V. K., Gao, Z., Lambkin, G. R., Xu, W., Winter, J. M., Marcus, S. L., et al. (2015). Comparison of 10,11-Dehydrocurvularin Polyketide Synthases from *Alternaria cinerariae* and Aspergillus terreus Highlights Key Structural Motifs. *Chembiochem* 16, 2479–2483. doi:10.1002/cbic.201500428.

Castro-Moretti, F. R., Gentzel, I. N., Mackey, D., and Alonso, A. P. (2020). Metabolomics as an Emerging Tool for the Study of Plant–Pathogen Interactions. *Metab*. 10. doi:10.3390/metabo10020052.

Depke, T., Thöming, J. G., Kordes, A., Häussler, S., and Brönstrup, M. (2020). Untargeted LC-MS Metabolomics Differentiates Between Virulent and Avirulent Clinical Strains of Pseudomonas aeruginosa. *Biomol*. 10. doi:10.3390/biom10071041.

Elosta, S., Gajdošová, D., Hégrová, B., & Havel, J. (2007). MALDI TOF mass spectrometry of selected mycotoxins in barley. *J Appl Biomed*, 5(1), 39-47. doi: 10.32725/jab.2007.006.

Feitosa-Junior, O. R., Stefanello, E., Zaini, P. A., Nascimento, R., Pierry, P. M., Dandekar, A. M., et al. (2019). Proteomic and Metabolomic Analyses of Xylella fastidiosa OMV -Enriched Fractions Reveal Association with Virulence Factors and Signaling Molecules of the DSF Family. *Phytopathology* 109, 1344–1353. doi:10.1094/PHYTO-03-19-0083-R.

Ghisalberti, E. L., and Rowland, C. Y. (1993). 6-chlorodehydrocurvularin, a new metabolite from cochliobolus spicifer. *J. Nat. Prod*. 56, 2175–2177. doi:10.1021/NP50102A022.

Gutiérrez, M., Theoduloz, C., Rodríguez, J., Lolas, M., and Schmeda-Hirschmann, G. (2005). Bioactive Metabolites from the Fungus Nectria galligena, the Main Apple Canker Agent in Chile. *J. Agric. Food Chem*. 53, 7701–7708. doi:10.1021/jf051021l.

Hassan, A. E. H. . (2007). Novel natural products from endophytic fungi of Egyptian medicinal plants : chemical and biological characterization. Available at: http://www.secheresse.info/spip.php?article5992 [Accessed August 28, 2021].

He, J., Wijeratne, E. M. K., Bashyal, B. P., Zhan, J., Seliga, C. J., Liu, M. X., et al. (2004). Cytotoxic and Other Metabolites of Aspergillus Inhabiting the Rhizosphere of Sonoran

Desert Plants. *J. Nat. Prod.* 67, 1985–1991. doi:10.1021/np040139d.

Hyeon, S.-B., Ozaki, A., Suzuki, A., and Tamura, S. (1976). Isolation of αβ-Dehydrocurvularin and β-Hydroxycurvularin from *Alternaria* tomato as Sporulation-suppressing Factors. *Agric. Biol. Chem.* 40, 1663–1664. doi:10.1080/00021369.1976.10862283.

Jeon, Y.-T., Ryu, K.-H., Kang, M.-K., Park, S.-H., Yun, H., QT, P., et al. (2010). Alternariol monomethyl ether and α, β-dehydrocurvularin from endophytic fungi *Alternaria* spp. Inhibit appressorium formation of Magnaporthe grisea. *J. Korean Soc. Appl. Biol. Chem.* 53, 39–42. doi:10.3839/JKSABC.2010.007.

Jiang, S.-J., Qiang, S., Zhu, Y.-Z., and Dong, Y.-F. (2008). Isolation and phytotoxicity of a metabolite from Curvularia eragrostidis and characterisation of its modes of action. *Ann. Appl. Biol.* 152, 103–111. doi:https://doi.org/10.1111/j.1744-7348.2007.00202.x.

Kobayashi, A., Hino, T., Yata, S., Itoh, T. J., Sato, H., and Kawazu, K. (1988). Agricultural and Biological Chemistry Unique Spindle Poisons, Curvularin and Its Derivatives, Isolated from Penicillium Species. 52, 3119–3123. doi:10.1080/00021369.1988.10869190.

Kusano, M., Nakagami, K., Fujioka, S., Kawano, T., Shimada, A., and Kimura, Y. (2003). Pratylenchus penetrans from the Fungus Aspergillus sp. *Biosci. Biotechnol. Biochem* 67, 1413–1416. Available at: https://academic.oup.com/bbb/article/67/6/1413/5944393 [Accessed August 11, 2021].

Kwak, M., Kang, K., and Wang, Y. (2019). Methods of Metabolite Identification Using MS/MS Data. *J. Comput. Inf. Syst.* doi:10.1080/08874417.2019.1681328.

Lai, S., Shizuri, Y., Yamamura, S., Kawai, K., Terada, Y., and Furukawa, H. (1989). Novel curvularin-type metabolites of a hybrid strain ME 0005 derived from penicillium citreo-viride B. IFO 6200 and 4692. *Tetrahedron Lett.* 30, 2241–2244. doi:10.1016/S0040-

4039(00)99658-8.

Lee, H. B., Patriarca, A., and Magan, N. (2015). *Alternaria* in Food: Ecophysiology, Mycotoxin Production and Toxicology. *Mycobiology* 43, 93–106. doi:10.5941/MYCO.2015.43.2.93.

Lou, J., Fu, L., Peng, Y., and Zhou, L. (2013). Metabolites from *Alternaria* fungi and their bioactivities. *Molecules* 18, 5891–5935. doi:10.3390/molecules18055891.

Meena, M., Gupta, S. K., Swapnil, P., Zehra, A., Dubey, M. K., and Upadhyay, R. S. (2017). *Alternaria* toxins: Potential virulence factors and genes related to pathogenesis. *Front. Microbiol.* 8, 1451. doi:10.3389/fmicb.2017.01451.

Meng, L.-H., Li, X.-M., Lv, C.-T., Li, C.-S., Xu, G.-M., Huang, C.-G., et al. (2013). Sulfur-Containing Cytotoxic Curvularin Macrolides from Penicillium sumatrense MA-92, a Fungus Obtained from the Rhizosphere of the Mangrove Lumnitzera racemosa. *J. Nat. Prod*. 76, 2145–2149. doi:10.1021/NP400614F.

Meronuck, R. A., Steele, J. A., Mirocha, C. J., and Christensen, C. M. (1972). Tenuazonic Acid, a Toxin Produced by *Alternaria alternata* . *Appl. Microbiol.* 23, 613–617. doi:10.1128/AM.23.3.613-617.1972.

Munro, H. D., Musgrave, O. C., and Templeton, R. (1967). Curvularin. Part V. The compound C16H18O5, αβ-dehydrocurvularin. *J. Chem. Soc. C Org.*, 947–948. doi:10.1039/J39670000947.

Musgrave, O. C. (1956). 828. Curvularin. Part I. Isolation and partial characterisation of a metabolic product from a new species of Curvularia. *J. Chem. Soc.*, 4301–4305. doi:10.1039/JR9560004301.

Robeson, D. J., Strobel, G. A., and Strange, R. N. (1985). The identification of a major phytotoxic component from *Alternaria* macrospora as αβ-dehydrocurvularin. *J. Nat. Prod*.

48, 139–141. Available at: https://pubs.acs.org/sharingguidelines [Accessed August 10, 2021].

Robeson, D., and Strobel, G. (1981). αβ-Dehydrocurvularin and Curvularin from *Alternaria* cinerariae. *Z. Naturforsch.* 36, 1081–1083. doi:10.1515/znc-1981-11-1231.

Steele, J. A., Durbin, R. D., Uchytil, T. F., and Rich, D. H. (1978). Tentoxin An uncompetitive inhibitor of lettuce chloroplast coupling factor 1. *Biochim. Biophys. Acta - Bioenerg.* 501, 72–82. doi:10.1016/0005-2728(78)90096-8.

Stone, G., and Sadowski, I. (1993). GAL4 is regulated by a glucose-responsive functional domain. *EMBO J.* 12, 1375–1385. Available at: https://pubmed.ncbi.nlm.nih.gov/8467796.

Vurro, M., Evidente, A., Andolfi, A., Chiara Zonno, M., Giordano, F., and Motta, A. (1998). Brefeldin A and α,β-dehydrocurvularin, two phytotoxins from *Alternaria zinniae*, a biocontrol agent of Xanthium occidentale. *Plant Sci.* 138, 67–79. doi:https://doi.org/10.1016/S0168-9452(98)00131-9.

Wang, M., Fu, H., Shen, X. X., Ruan, R., Pun, N., Xu, J., et al. (2017). Genomic features and evolution of the conditionally dispensable chromosome in the tangerine pathotype of *Alternaria alternata*. *bioRxiv*. doi:10.1101/207746.

Wang, Y., Kora, G., Bowen, B. P., and Pan, C. (2014). MIDAS: A Database-Searching Algorithm for Metabolite Identification in Metabolomics. *Anal. Chem.* 86, 9496–9503. doi:10.1021/ac5014783.

Witte, T., Harris, L., Nguyen, H., Hermans, A., Johnston, A., Sproule, A., et al. (2020). Apicidin biosynthesis is linked to accessory chromosomes in *Fusarium Poae* isolates. *BMC Genomics* Preprint. doi:10.21203/rs.3.rs-116075/v1.

Witte, T. E., Villeneuve, N., Boddy, C. N., and Overy, D. P. (2021). Accessory Chromosome-

Acquired Secondary Metabolism in Plant Pathogenic Fungi: The Evolution of Biotrophs Into Host-Specific Pathogens . *Front. Microbiol.* 12, 700. Available at: https://www.frontiersin.org/article/10.3389/fmicb.2021.664276.

Xiang, C., Liu, Y., Liu, S.-M., Huang, Y.-F., Kong, L.-A., Peng, H., et al. (2020). αβ-Dehydrocurvularin isolated from the fungus *Aspergillus welwitschiae* effectively inhibited the behaviour and development of the root-knot nematode Meloidogyne graminicola in rice roots. *BMC Microbiol. 2020 201* 20, 1–10. doi:10.1186/S12866-020-01738-2.

Xie, L. W., Ouyang, Y. C., Zou, K., Wang, G. H., Chen, M. J., Sun, H. M., et al. (2009). Isolation and difference in anti-Staphylococcus aureus bioactivity of curvularin derivates from fungus Eupenicillium sp. *Appl. Biochem. Biotechnol.* 159, 284–293. doi:10.1007/S12010-009-8591-2.

Xu, Y., Espinosa-Artiles, P., Schubert, V., Xu, Y., Zhang, W., Lin, M., et al. (2013). Characterization of the biosynthetic genes for 10,11-dehydrocurvularin, a heat shock response-modulating anticancer fungal polyketide from Aspergillus terreus. *Appl. Environ. Microbiol.* 79, 2038–2047. doi:10.1128/AEM.03334-12.

Ząbek, A., Klimek-Ochab, M., Jawień, E. et al. (2017). Biodiversity in targeted metabolomics analysis of filamentous fungal pathogens by 1H NMR-based studies. *World J Microbiol Biotechnol* 33, 132. https://doi.org/10.1007/s11274-017-2285-7

Zhan, J., Wijeratne, E. M. K., Seliga, C. J., Zhang, J., Pierson, E. E., PIERSON III, L. S., et al. (2004). A New Anthraquinone and Cytotoxic Curvularins of a Penicillium sp. from the Rhizosphere of Fallugia paradoxa of the Sonoran Desert. *J. Antibiot. (Tokyo).* 57, 341–344. doi:10.7164/ANTIBIOTICS.57.341.

3.7 Appendix B

3.7.1 Chemicals and Solvents

Table 3.4. A list of the chemicals and solvents, along with their manufacturer's, utilized to perform the experimentation conducted in Chapter 3 of the present thesis.

Chemical	Manufacturer
Ethanol (EtOH)	Thermo-Fisher Scientific
Ethyl acetate (HPLC-grade)	EMD Millipore
Methanol (LCMS-grade)	VWR Chemicals
Acetonitrile (Optima, LCMS-grade)	Thermo-Fisher Scientific
Water (Optima, LCMS-grade)	Thermo-Fisher Scientific
Hexanes	Thermo-Fisher Scientific
Nanopure water	Thermo-Fisher Scientific, Barnstead Nanopure
Sucrose	Anachemia
Corn meal	Unico
Yeast extract	Oxoid
$MgSO_4 \cdot 7H_2O$	BDH
Agar	VWR
ACRODISC 0.45µm PFTE filters	Pall Corporation
Formic acid	Sigma-Aldrich
Methanol- d_4	Sigma-Aldrich
Chloroform	uOttawa laboratory stores
TLC plate (250 µm thick, 20x20cm plate)	Merck
Acid-washed sea sand	uOttawa laboratory stores

Chapter 4

4 Identifying and Characterizing the Biosynthetic Gene Clusters Responsible for Selected Natural Products

4.1 Introduction

Over the past few decades, innovations in science and technology have fueled significant progress in ways previously unimaginable. One of which has been in understanding how members of the largely plant-pathogen dominated class of Dothideomycetes, which includes genera like *Alternaria*, cause disease at a molecular level (Muria-Gonzalez *et al.*, 2015). This can largely be attributed to the modern genomic era, transforming natural product research into an interdisciplinary arena for tackling complex questions like the origins, genetic regulation and biosynthesis of natural products. The tools at our disposal (like the increasing availability of telomere to telomere genome sequences) have not only revealed the remarkable processes responsible for SMs, but also that the capacity of fungi to produce SMs has been significantly underestimated (Wiemann and Keller, 2014).

While the structural diversity of the fungal SMs is vast, they can be broadly grouped into four main categories based on their biosynthetic origin: polyketides made by polyketide synthases (PKSs), non-ribosomal peptides (NRPs) made by non-ribosomal peptide synthases (NRPSs), terpenoids made by prenyltransferases (PTs)/terpene synthases (TSs) and tryptophan derivates made by dimethylally tryptophan synthases (DMATSs) (Wiemann and Keller, 2014; Muria-Gonzalez *et al.*, 2015). The conserved sequence features within the respective families of core

enzymes, can be exploited to allow for the rapid (and often automated) identification of genes encoding PKSs, NRPSs, PTs, TSs and DMATSs within fungal genomes (Muria-Gonzalez *et al.*, 2015). The fact that genes involved in fungal SM biosynthesis are often clustered together on chromosomes (in biosynthetic gene clusters (BGCs)) further facilitates the identification of these enzymes (Keller *et al.*, 2005; McGary *et al.*, 2013; Muria-Gonzalez *et al.*, 2015). Although BGCs can be readily identified from sequenced fungi, the number of characterizable SMs produced by sequenced fungi remains limited for a variety of reasons (Muria-Gonzalez *et al.*, 2015). This is where the "multi-omics" strategy of metabolomics and genomics can be of significant benefit.

As discussed in **Chapter 1**, combining genomics and metabolomics is a powerful way to screen populations of fungi for SM production. Modern genome sequencing allows for the experimenter to detect nearly all the SM BGCs in a genome, even if they are silent under laboratory conditions (Witte et al., 2021). Metabolomics allows the experimenter to analyze the chemical extracts of fungal strains to characterize and compare patterns in SM production observed within a population, to identify unique patterns attributable to sub-sets within the population (Witte et al., 2021). This "multi-omics" strategy is currently being applied in studies of agriculturally relevant crop pathogens, has proved successful in the characterization of the apicidins in *F. poae*, and has been utilized to reveal key metabolic pathways and regulators of Alzheimer's disease (Witte et al., 2020; 2021; Horgusluoglu et al., 2021).

With high-quality genome sequences, there is a wealth of invaluable information that can help form a better understanding of the biological functions that a given organism carries out. Given

the immense size of genomes, manual identification of meaningful sequence information from genome data is not feasible. Instead, computer-based techniques are required to efficiently and accurately analyze the data. This is accomplished through the application of several signal processing models and algorithms (Yoon, 2009). Among those that been applied in biological sequence analysis, hidden Markov models (HMMs) is one of the most popular as it is effective in modeling the correlations between adjacent variable, domains, or events (Yoon, 2009). HMM is a statistical model built from the alignment of protein sequences known to have the same function.

In the context of the widely-used genome mining application, antiSMASH, a profile HMM is utilized to predict the location and architecture of the BGCs in a given genome (Blin et al., 2021). Profile HMMs are statistical models that convert information from multiple sequence alignment into a set of probability values that denote position-specific variation levels in all members of evolutionarily related sequences (Reyes et al., 2017). In this way, they are able to capture a wide array of variation and have higher sensitivity that conventional similarity methods like BLAST for the detection of remote homologous sequences (Reyes et al., 2017).

Most fungal polyketides are biosynthesized using a sole iterative polyketide synthase (iPKS) enzyme (Wang et al., 2020). Contrary to the norm, previous characterization of the DHC BGC in an *A. terreus* isolate (strain AH-02-30-F7) and an *A. cinerariae* isolate has proved it involves the use of 2 separate PKSs (Xu et al., 2013; Cochrane et al., 2015). Specifically, it involves the use of a highly-reducing iterative PKS (hrPKS) and a non-reducing iterative PKS (nrPKS), in a mechanism very similar to that employed for RAL biosynthesis.

All fungal PKSs are comprised of a single ketoacyl synthase (KS), acyltransferase (AT), and acyl carrier protein (ACP) core domains. They iteratively build the carbon scaffold of the polyketide by recursive, decarboxylate Claisen condensations of malonyl-coenzyme A (CoA) precursors (Xu et al., 2013). For RAL biosynthesis, the hrPKS is comprised of a ketoacyl reductase (KR), dehydratase (DH), and enoyl reductase (ER) domains that may reduce the growing polyketide chain to an alkane, alkene, or alcohol after each condensation step (Crawford and Townsend, 2010; Chooi and Tang, 2012). The activity of the hrPKS thereby results in a reduced polyketide chain (Crawford and Townsend, 2010; Chooi and Tang, 2012).

In RAL biosynthesis, the polyketide chain then gets transferred to the nrPKS by the starter AT domain (SAT) of the nrPKS, where it undergoes further elongation (Foulke-Abel and Townsend, 2012; Xu et al., 2013). The product template (PT) domain of the nrPKS is responsible for ring closure and, in the context of RAL biosynthesis, this means a C2-C7 aldol condensation (yielding the resorcinol carboxylate moiety) (Crawford et al., 2008, 2009; Li et al., 2010; Xu et al., 2013). The subsequent thioesterase (TE) domain is then responsible for bridging the macrolactone ring (Wang et al., 2009). As such, the TE domain of the nrPKS is where the RAL is off-loaded/released (Xu et al., 2013).

Characterization of the DHC BGC by Xu *et al.* (2013) (in *A. terreus*) revealed that the biosynthesis of DHC is carried out in much the same way as RALs. The gene they predicted to be the hrPKS involved in DHC biosynthesis (AtCURS1) shared end-to-end similarity and had identical domain composition to the characterized RAL hrPKS. Additionally, the gene they

predicted to be the nrPKS involved in DHC biosynthesis (AtCURS2) featured identical domain composition to that of the nrPKS characterized in RAL biosynthesis.

A subsequent study by Cochrane *et al*. (2015), characterized the DHC BGC in an *A. cinerariae* isolate as shown in **Figure 4.1.** Comparison of the BGC in the *A. cinerariae* isolate to that which was characterized in the *A. terreus*, revealed that the hrPKS (Dhc3) and nrPKS (Dhc5) shared 75% and 77% sequence identity with CurS1 and CurS2, respectively (Cochrane et al., 2015). The domains of the nrPKS of *A. cinerariae* and *A. terreus* are identical, but the hrPKS domains differ. Specifically, while as the hrPKS domains of *atcurs1* are KS-AT-DH-KR_s-ER-KR_c-ACP, the hrPKS domains of *dhc3* are KS-MAT-DH-PMT-PKR-ER-KR_c-ACP (where KS is ketosynthase, AT is acyltransferase, DH is dehydrogenase, KR_s is ketoreductase structural subdomain, ER is enoylreductase, KR_c is ketoreductase catalytic domain, ACP is acyl carrier protein, MAT is malonyl-CoA:ACP acyltransferase, PMT is pseudo C-methyltransferase, and PKR is pseudo ketoreductase) (Xu et al., 2013; Cochrane et al., 2015). Despite these differences, they both produce the same product(s).

The curvularins have been isolated from fungal isolates of several genera, including *Alternaria*, *Aspergillus, Cochliobolus, Curvularia, Nectria,* and *Penicillium* spp. (Musgrave, 1956; Hyeon et al., 1976; Robeson et al., 1985; Ghisalberti and Rowland, 1993; He et al., 2004; Zhan et al., 2004; Gutiérrez et al., 2005). The *Alternaria spp*. that have been found to produce the curvularins currently includes *Alternaria cinerariae, Alternaria zinniae, Alternaria macrospora*, and *Alternaria sp*. isolate from the leaves of *Acaia mangium*, and *Alternaria tomato* (which would now typically be referred to as an *Alternaria alternata* tomato pathotype isolate) (Hyeon

et al., 1976; Robeson et al., 1985; Vurro et al., 1998; Jeon et al., 2010; Cochrane et al., 2015). The results of the previous chapter suggest that *A. arborescens* should now be included amongst the curvularins-producing *Alternaria* species'.

As discussed in the previous chapter, the curvularins are a class of polyketides that have been reported to exhibit phytotoxic, antimicrobial, antitumoral, nematicidal, and phytotoxic activities (Hyeon et al., 1976; Caputo and Viola, 1977; Vurro et al., 1998; Kusano et al., 2003; Gutiérrez et al., 2005; Xiang et al., 2020). In eukaryotic cells, DHC has been reported to be a strong activator of the heat shock response (Santagata et al., 2012). Introduction of DHC to the cell causes an overexpression of the Heat Shock Factor 1 (HSF1) which, in turn, has shown to be an inhibitor of various cancer cell lines in vitro (Santagata et al., 2012). It's mechanism of action in plant tissue, bacteria, and fungi, however, remains poorly understood.

The purpose of the current chapter was to identify and characterize the BGCs in the genomes of the DET2035, KAS5321, and KAS5743 isolates that correspond with the structural motifs of the metabolites identified in Chapters 3 and to assess the relatedness of this BGC with other fungi. Towards accomplishing this task, the whole genome sequences of the two *A. arborescens* isolates (DET2035 and KAS5321) and *A. alternata* isolate (KAS5743) were obtained. The whole genome sequences were subsequently mined through the use of fungal antiSMASH (v6.0.1) to determine the function and location of the BGCs in their genomes. Mining was subsequently biased towards retrieving the DHC BGC. It was hypothesized that associated BGCs of the unique set of metabolites is in the genome of the isolates from which the curvularins were

detected, and that the origin of the BGC can be implicated as being shared amongst different fungal plant pathogens.

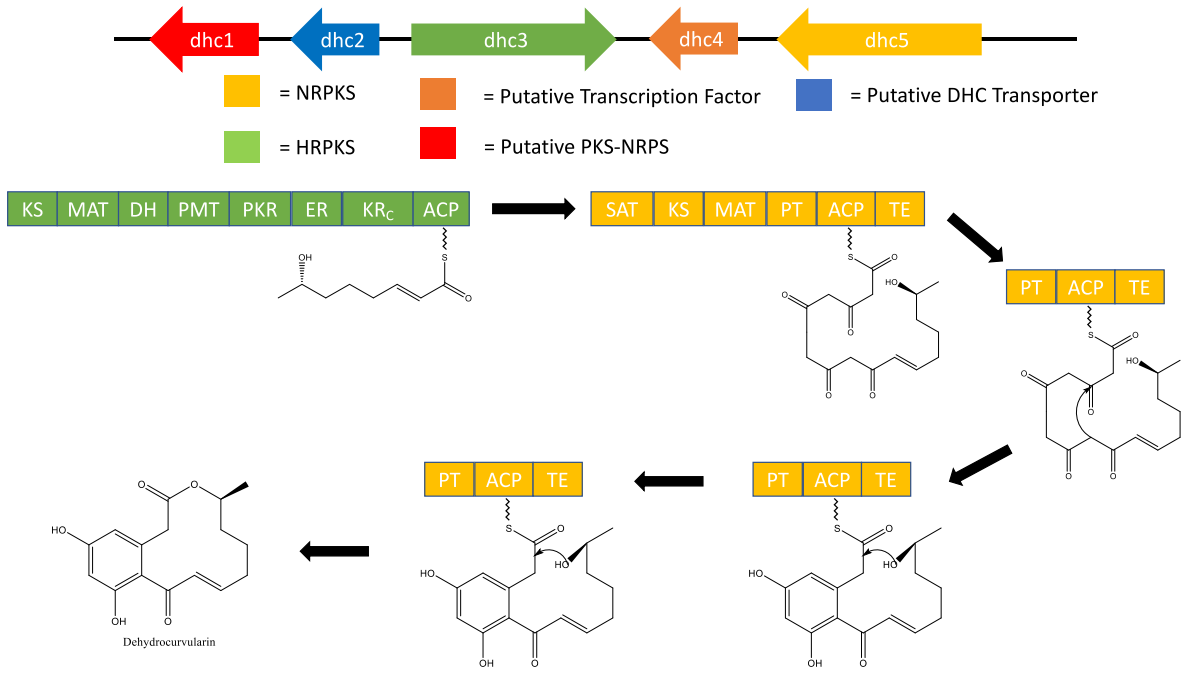

Figure 4.1. The dehydrocurvularin (DHC) biosynthetic gene cluster (BGC) as it is reported in its characterization in the *A. cinerariae* isolate, by Cochrane et al. (2015).

4.2 Materials and Methods

4.2.1 Fungal Strains

Based on the results of the previous chapters of the present thesis, DET2035, KAS5321, and KAS5743 were selected for whole genome sequencing analysis.

4.2.2 Whole Genome Sequencing

The whole genome sequence data for the 3 selected strains, were obtained by a collaborating lab group. DET2035 was subjected to Nanopore (Oxford Nanopore Technologies) and Illumina (Ambry Genetics) genome sequencing. The combination of both sequencing technologies allowed for the sequencing data to be assembled into long contigs due to the long-read genome sequence data from the Nanopore sequencing and error corrected with the more accurate Illumina short-read sequences. The whole genome sequences of KAS5321 and KAS5743 were obtained through Illumina whole genome sequencing.

4.2.3 Genome Mining

The profile of BGCs encoding SMs in the 3 *Alternaria* strains were predicted and annotated using fungiSMASH v6.0.1, which identifies BGCs using a profile Hidden Markov Model (Blin et al., 2021). The profile of BGCs was then mined for the DHC BGC. Once the BGC had been located in each of the genomes, the genbank (.gbk) file for the region was downloaded from fungiSMASH so it could be utilized by other programs.

4.2.4 Sequence Similarity Networks

After downloading the annotated DHC BGC from each of the 3 *Alternaria* strains, the amino acid sequence for the predicted hrPKS and nrPKS were retrieved and utilized as query sequences in the EFI – Enzyme Similarity Tool (EST) (Gerlt et al., 2015; Gerlt, 2017; Zallot et al., 2018, 2019). The parameters utilized in the EFI – EST analysis were where the maximum BLAST sequences was set to 1000, the maximum length was set to 50,000, the filter type set to e-value, and the filter value set to 50. In order to generate local genome neighbourhood diagrams, the EFI – GNT (genome neighbourhood tool) was utilized (Gerlt et al., 2015; Gerlt, 2017; Zallot et al., 2018, 2019). The parameters utilized for this analysis were that the neighbourhood size was set to 10 and the minimal co-occurrence percentage lower limit was set to 20. These settings were utilized for the nrPKS and hrPKS of the DHC BGC in all three *Alternaria* strains utilized in the present chapter.

The outputted sequence similarity network was then imported into the open-source bioinformatic software, Cytoscape (downloaded from https://cytoscape.org/). Cytoscape allowed for the data to be visualized as a network of nodes and edges through use of the yFiles organic layout (yWorks GmbH is the maker of yFiles). The nodes represented the genes which were a match to the query sequence, and the edges depicted the similarity relationship between any two nodes. Once visualized as a network, the specificity of the edges was set such that only the nodes with a relationship of 85% identity matches or greater would be connected by edges. If the relationship between any two nodes was lesser than 85%, the edge connecting them was deleted from the network.

4.3 Results

Figure 4.2. Screenshot of the antiSMASH v6.0.1 user interface, whereby the T1PKS region that denotes the DHC BGC is being highlighted, in a utilized *A. arborescens* isolate (DET2035). The non-reducing PKS (nrPKS) and highly-reducing PKS (hrPKS) of the BGC are denoted by **A**, and **B,** respectively**.** The domains that comprise the genes of the gene cluster (the nrPKS and hrPKS) are denoted by **C** (with the nrPKS being FUN_011019 and the hrPKS being FUN_011021).

The antiSMASH bioinformatics platform was used to identify the biosynthetic gene clusters in the genome of an *A. arborescens* isolate (DET2035) (**Figure 4.2**). The same technique of genome mining was applied another *A. arborescens* isolate (KAS5321) and an *A. alternata* isolate (KAS5743) (**Figure 4.8** and **Figure 4.9** of **Appendix C**, respectively). The three isolates were selected on the basis of the results of the previous chapters, which showed they each produced DHC and its analogues. Interestingly, the gene cluster encoding DHC-biosynthesis was found in all three isolates, and the organization of the clusters was similar to one another. Comparison of the organization of the DHC clusters is shown in **Figure 4.3**.

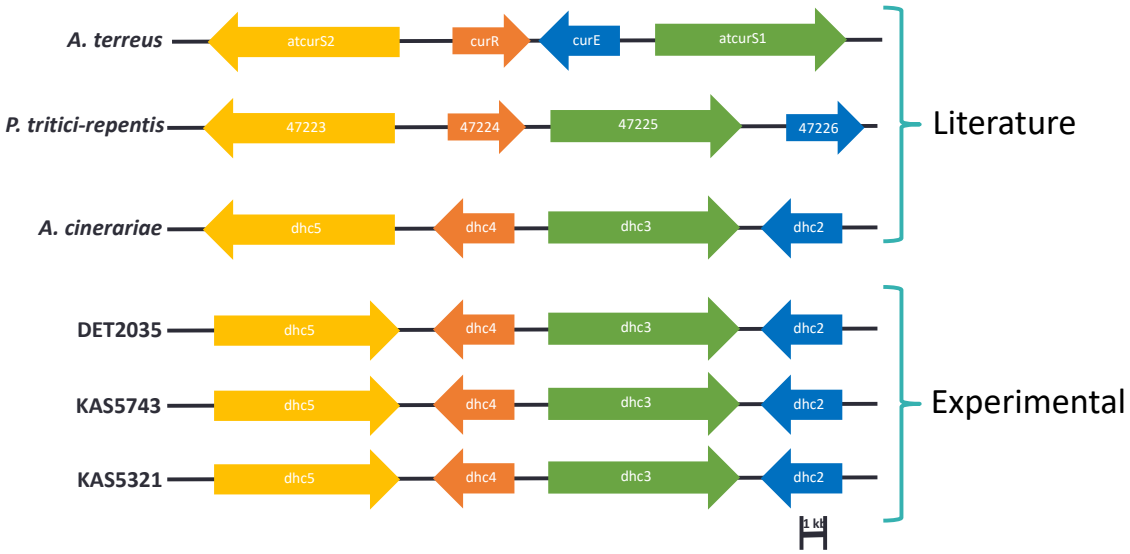

Figure 4.3. Comparison of the organization of the genes comprising the DHC BGC in the three isolates of study to literature.

Comparison of the DHC BGCs between the three *Alternaria* isolates utilized in the present chapter, may be visualized with **Figure 4.3**, above. The DHC BGC predicted in all 3 isolates contains 4 genes. Interestingly, the organization of the genes in the clusters is similar between all three isolates, despite that two were *A. arborescens* isolates and one was an *A. alternata* isolate.

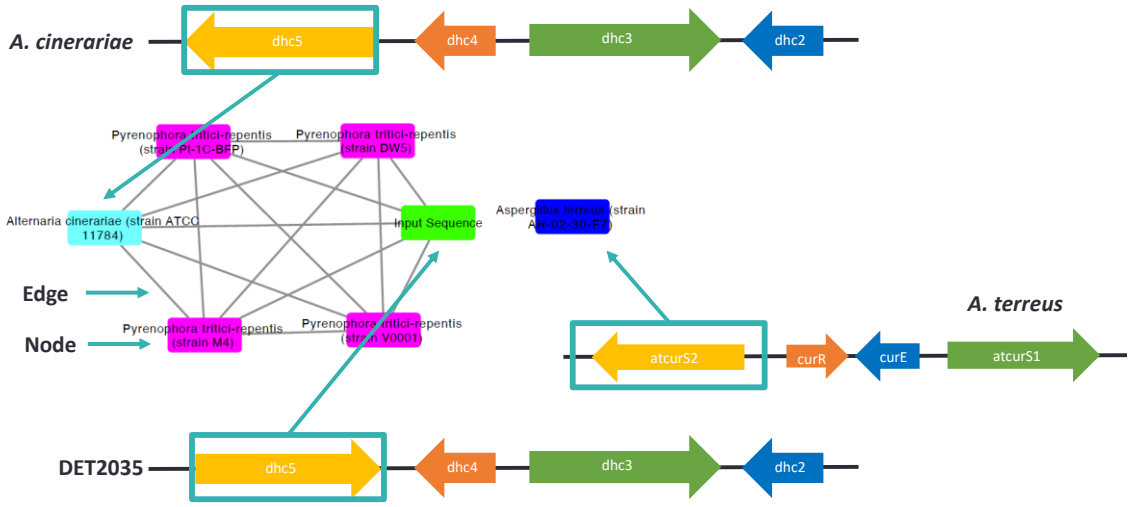

Figure 4.4. The sequence similarity network generated from the use of the nrPKS of the DHC BGC of DET2035 as the query sequence, with 10 nodes represented. Each network (or cluster of nodes) represents the nodes that could be connected with highest similarity to one another (with > 85% identity and >= 1150 alignment score) (except *A. terreus*). Nodes highlighted in bright blue denote those from *Alternaria*, pink denotes those from *Pyrenophora*, dark blue denotes those from *Aspergillus*. The input sequence is denoted by the bright green node.

A sequence similarity network (SSN) from the nrPKS of the DHC BGC in DET2035 is shown in **Figure 4.4**. An SSN creates clusters of nodes based on sequence similarity and each cluster is to therefore represents a common function. The singleton node representing the nrPKS of the DHC BGC of *A. terreus* was kept in the SSN as it is a node with a known function. Contrary to what would be expected based on its known role in DHC biosynthesis, it did not form an edge with the nrPKS of DET2035 or with the *A. cinerariae*. Interestingly, reduction of the edge requirement to just 70% identity matches was still too high a threshold for *A. terreus* to form an edge with the network containing *A. cinerariae* (and the input sequence).

Inspection of **Figure 4.4** revealed that there was a network of closely related nodes which was comprised of homologs from *Pyrenophora* and *Alternaria*, as well as the input sequence (DET2035). This is consistent with the data from **Table 4.1** of **Appendix C**, where *A. cinerariae* and *Pyrenophora tritici-repentis* are the two highest matches to the DET2035 nrPKS. Additionally, this would be expected as the known function of the *A. cinerariae* node is as an nrPKS involved in DHC biosynthesis. The local genome neighbourhoods of the PKS's comprising this network are shown in **Figure 4.5. Figure 4.5** shows that the DHC BGC is evidently conserved across *P. tritici-repentis*, *A. cinerariae*, and the *A. arborescens* utilized for this study (DET2035), sharing an identical organization and synteny with the characterized *A. cinerariae* DHC BGC. The hrPKS and nrPKS of each of the clusters shown in **Figure 4.5** were inputted into BLASTp to assess if they were of the DHC BGC. In each instance, they showed a high level of similarity to the nrPKS of *A. cinerariae* (>= 85% percent identity and 100% query coverage). This all goes to prove that the DHC BGC has been detected in DET2035. Contrary to what would be expected based on **Table 7** of **Appendix C**, a node to represent *A. solani* did not show up as a match in the sequence similarity network. Analysis of the nrPKS from KAS5321 and KAS5743 show identical results as those from DET2035 (see appendix **Figure 4.10** and **Figure 4.12**).

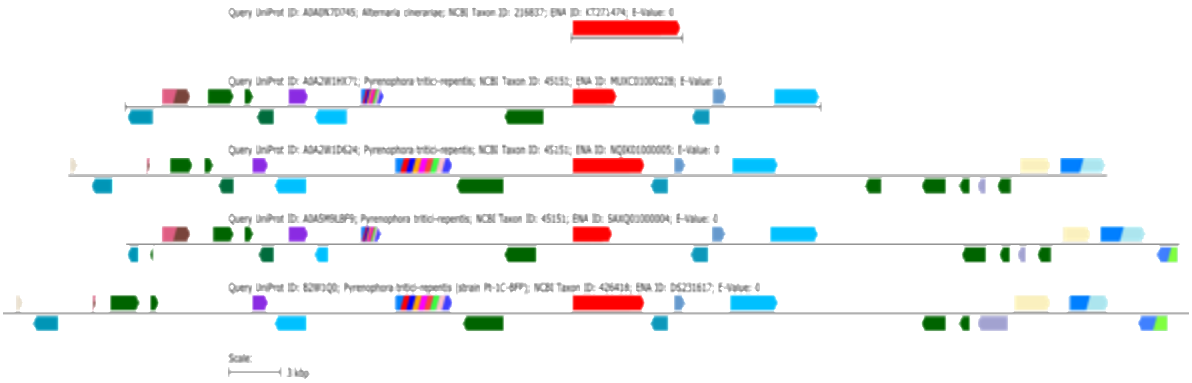

Figure 4.5. The local genome neighbourhood that represents the network of *Pyrenophora*, *A. cinerariae*, and the input sequence (DET2035). When generating a genome neighbourhood, the input sequence for the SSN is not retrievable. The nrPKS is represented in red, and the hrPKS is represented in multi-color. The scale-bar represents 3 kbp.

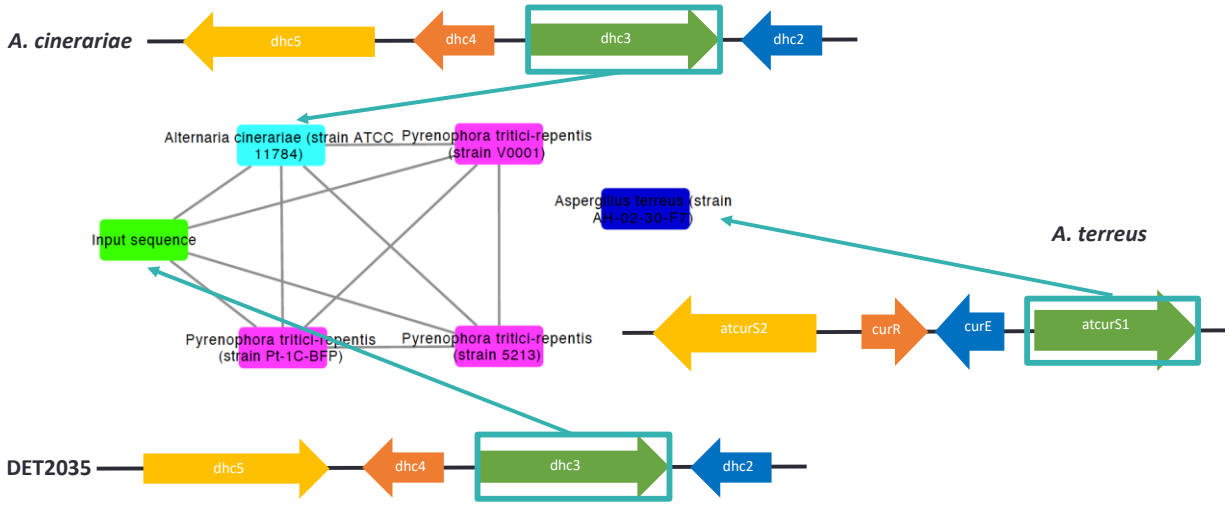

Figure 4.6. The sequence similarity network generated from the use of the hrPKS of the DHC BGC of DET2035 as the query sequence, with 7 nodes represented. Each network (or cluster of nodes) represents the nodes that could be connected with highest similarity to one another (with > 85% identity and >= 1150 alignment score). Nodes highlighted in bright blue denote those from *Alternaria*, pink denotes those from *Pyrenophora*, dark blue denotes those from *Aspergillus*, and the input sequence is denoted by the bright green node.

The SSN from the hrPKS of the DHC BGC in DET2035 is shown in **Figure 4.6**. The node of the hrPKS of the DHC BGC of *A. terreus* is represented by the node labelled L7X8J4 (**Figure 4.6**). Though it shows an edge relationship with another *Aspergillus spp.*, contrary to what would be expected based on its known role in DHC biosynthesis, it did not form an edge with the hrPKS of DET2035, *A. cinerariae* (A0A0N7D4P6). Again, reduction of the edge requirement to just 70% identity matches did not result in the *A. terreus* forming an edge with the network containing *A. cinerariae*.

Inspection of **Figure 4.6** revealed that, similar to **Figure 4.4**, the *A. cinerariae*, input sequence, and several *P. tritic-repentis* nodes cluster into a distinct network. **Figure 4.7** shows that this clustering pattern is the result of the DHC BGC being conserved amongst these isolates. Analysis of the hrPKS from KAS5321 and KAS5743 show identical results as those from DET2035 (see **Appendix C**, **Figure 4.11** and **Figure 4.13**).

Figure 4.7. The local genome neighbourhood that represents the network of *Pyrenophora*, *A. cinerariae*, and the input sequence (DET2035). When generating a genome neighbourhood, the input sequence for the SSN is not retrievable. The hrPKS is represented in red, and the nrPKS is represented in multi-color. The scale-bar represents 3 kbp.

4.4 Discussion

The purpose of the current chapter was to identify and characterize the BGCs in the genomes of *A. sect. A. alternata* isolates that correspond with the structural motifs of the metabolites identified in Chapters 3 and to assess the relatedness of this BGC with other fungi. Towards accomplishing this task, the whole genome sequences of the two *A. arborescens* isolates (DET2035 and KAS5321) and an *A. alternata* isolate (KAS5743) were obtained. The whole genome sequences were subsequently mined through the use of fungal antiSMASH (v6.0.1), to determine the function and location of the BGCs in their genomes. Mining was subsequently biased towards retrieving the DHC BGC. It was hypothesized that the family of compounds/analogs obtained in **Chapter 3** are produced by the dehydrocurvularin (DHC) BGC, that is shared by DET2035, KAS5321, and KAS5743. Additionally, that this BGC can be implicated as being shared amongst different fungal plant pathogens.

Comparison between the DHC BGC predicted from DET2035, KAS5321, and KAS5743 revealed identical organizational patterns. This was expected as DET2035 and KAS5321 are two *A. arborescens* isolates, while KAS5743 is an *A. alternata* isolate. As such, it was expected that the DHC BGC and, a large portion of the genome of the three isolates would be similar to one another based on that molecular techniques have not been able to resolve the *A. alternata* and *A. arborescens* species-complexes from one another. Similarity between the 2 genomes was also to be expected based on that KAS5321 and DET2035 consistently clustered next to one another according to the profile of SMs they produced (in the metabolomics work of **Chapter 2)**.

A study by conducted by Xu et al. (2013) characterized the DHC BGC an *A. terreus* isolate (AH-02-30-F7). In this isolate, they reported that the biosynthesis of DHC is carried out by the combinatorial work of an nrPKS (AtCURS1) and a hrPKS (AtCURS2) (Xu et al., 2013). Shortly thereafter, the DHC BGC was characterized in an *A. cinerariae* (ATTC 11784) isolate by Cochrane et al. (2015). To date, these are to only two reports where the DHC BGC has been well characterized. While as *P. tritici-repentis* has not been described to biosynthesize curvularin, the DHC BGC has been found in its genome.

Comparison of the organization of the DHC BGC observed in the three sequenced *Alternaria* to that reported in a *A. cinerariae,* revealed a similar gene cluster organization (Cochrane et al., 2015). The domains comprising the hrPKS reported for *A. cinerariae,* however, were KS-MAT-DH-PMT-PKR-ER-KR_c-ACP (Cochrane et al., 2015). Comparison to domains of the hrPKS of the sequenced strains (KS-AT-DH-ER-KR-ACP) reveals that a PMT and PKR were not present. Running the hrPKS of *A. cinerariae* and all three sequenced strains through pfam, however, did not reveal the presence of these pseudo regions. As the SSN revealed that the 3 sequenced strains had a hrPKS that was highly identical to *A. cinerariae,* perhaps there is not an HMM for predicting PMT and PKR.

Comparison of the DHC BGC of the three sequence strains to *A. terreus* revealed that the organization of the gene cluster only differed slightly from literature (Xu et al., 2013). Specifically, the DHC BGC is reported in *A. terreus* to be comprised of a hrPKS (*atcurS2*), putative DHC transporter (*curE*), nrPKS (*atcurS1*), and putative transcription factor (*curR*) (Xu et al., 2013). The *Alternaria* strains sequenced for this study were observed to contain all 4 of

these genes, though putative DHC transporter was not located in-between the nrPKS and hrPKS. Additionally, the domains composing the hrPKS of the DHC BGC in *A. terreus* are reported to be KS-AT-DH-KR_s-ER-KR_c-ACP (Xu et al., 2013). This differs from the domains predicted by fungiSMASH. As such, the hrPKS of *A. terreus* and DET2035 were inputted into pfam. Neither of them revealed the presence of a KR_s and KR_c domain. As such, this may be another limitation of utilizing HMMs to predict the function of genes, where perhaps they do not have the ability to recognize KR_s and KR_c domains. These comparisons show that the DET2035, KAS5321, and KAS5743 DHC BGCs are homologous to the known DHC BGCs, but more so to *A. cinerariae* than *A. terreus*.

For a higher resolution analysis of BGC similarity, SSNs were generated. The SSNs represents the comparison of the amino acid sequence of the nrPKS and/or hrPKS to what is publicly available in the NIH/BLAST database. The limit of matches to consider was set to 1000 for the purpose of network generation. By then filtering the network to only map the edges which conferred a percent identity match of 85% or greater and an alignment score of 1150 or greater (with the maximum obtained in the analysis being 1400), only closely related genetic relations were shown.

Across all the sequence similarity networks presented in the current chapter, one may note that the input sequence (either the hrPKS or nrPKS of DET2035, KAS5321, or KAS5743) consistently formed a network with *A. cinerariae* and 3 or 4 *Pyrenophora* nodes. This suggested that the DHC BGC is highly conserved across *P. tritici-repentis*, *A. cinerariae*, and the *A. alternata* and *A. arborescens* spp. studied. It is believed that the curvularins are nHSTs. They are

not strictly required for pathogenicity but improve the pathogenic potential of a species in some non-host-specific way. As such, conservation of the BGC would be expected due to positive selective pressures. Interestingly, this is not the case when considering the *A. terreus* in which the DHC BGC was first characterized. A node representing the hrPKS or nrPKS of the DHC BGC in *A. terreus* did not form a network with an *Alternaria*-containing network, suggesting that they do not share a high level of sequence similarity. Presumably this is related to the ancient divergence of *Aspergillus* and *Alternaria* from their common Ascomycota ancestor (Wiese et al., 2011; Cochrane et al., 2015).

The tight network formed to model the DHC BGC of the 3 experimental *Alternaria* strains, *A. cinerariae,* and 3 or 4 *P. tritici-repentis* isolates suggests a genetic relationship that should be investigated. *P. tritici-repentis, A. alternata,* and *A. arborescens* are all common fungal pathogens of the triticale plant. Whole genome comparative analyses have been performed on *P. tritici-repentis,* where a high level of intraspecific genome plasticity was revealed in this species (Moolhuijizen et al., 2018). Additionally, a high level of genome plasticity is expected in *A. alternata,* being that they are currently believed to have acquired whole ACs through horizontal gene transfer (HGT) events. As such, it stands to reason that the DHC BGC was transferred from a *P. tritici-repentis* isolate to an *A. alternata* or *A. arborescens* by an HGT event, while they were both occupying a shared triticale host. Alternatively, perhaps *P. tritici-repentis, A. cinerariae, A. alternata, A. arborescens,* and *A. terreus* all share a distant common ancestor that has lost the ability to produce the curvularins over time. This would be very interesting at it would completely oppose the traditional view of evolution by positive selection. Conversely, the

lack of sequence similarity between *A. terreus* and the network of *A. cinerariae*, *P. tritici-repentis*, and the 3 *Alternaria* sequenced could indicate a case of convergent evolution.

Of course, the insights of the conducted analyses are not without their limitations. Firstly, the analysis was limited since it only mapped the similarity of the query sequence to that which is currently publicly available. As such, while as the DHC BGC is clearly conserved across the 3 *Alternaria* strains sequenced in this study, the *A. cinerariae*, and *P. tritici-repentis*, it could well be that the BGC is conserved across many other pathogenic fungi as well. Conversely, a DHC BGC that is more similar to that of *A. terreus* could exist (beyond the *A. novofumigatus* that occasionally showed similarity to the hrPKS of *A. terreus* in the generated SSNs) but is not currently in a publicly-available genome database.

Another limitation could manifest in that the selected percent identity requirement was set to 85% or greater and the alignment score set to 1150 or greater. Any singleton nodes and edges displaying any less than this level similarity to any other neighbouring node in the dataset, were deleted from the network (apart from the node representing the *A. terreus* characterized by Xu et al. (2013)). Additionally, only the most informative, non-singleton or paired networks were kept in the networks presented in the present chapter. The rationale for selection of this high threshold was in accordance with the goal of this study. The utilized approach successfully accomplished the experimental goal; however, it was at the loss of some nodes/relationships being modeled.

The function of DHC is not currently well understood and the specifics of how it induces effects have yet to be elucidated. It therefore uncertain why various fungi actively maintain the DHC

BGC. The SSNs make it evident that the DHC BGC is likely to play a role in virulence, hence why it is conserved across several fungal genera. While as DHC has been observed to arrest the cell cycle in cells by causing a microtubule assembly disturbance, this could have just been an artifact of its activity in the cell or a by-product of its activity (Kobayashi et al., 1988). Future work should be towards elucidating the bioactivity of DHC and its analogues. Bioactivity studies could be performed in planta, allowing for the role of the curvularins in virulence to elucidated.

4.5 Conclusion

The purpose of the current chapter was to identify and characterize the cluster of genes responsible for the production of the curvularins elucidated in **Chapter 3.** Towards accomplishing this task, the whole genomes of the three strains observed to produce the curvularins (DET2035, KAS5321, and KAS5743) were obtained and subsequently mined for the DHC BGC. It was hypothesized that the BGC would be located in all three strains, and that the BGC would be comparable to that of other fungal pathogens. The results obtained in the present chapter served to confirm the hypothesis and provide further evidence for the findings of the previous chapters. Specifically, the isolation and purification of DHC served to support that the metabolomics analyses were performed correctly in **Chapter 2**. Subsequently, the results of the present chapter served to confirm why DHC and its analogues were so integral to the metabolomic analyses, and why DHC was able to be isolated and purified (in **Chapters 2** and **3**, respectively).

The results of this experiment denote the first characterization of the DHC BGC in an *A. alternata* or *A. arborescens* isolate. Additionally, the generated SSNs suggest an interesting evolutionary history of the DHC BGC. Perhaps the genus *Alternaria* acquired the DHC BGC from *P. tritici-repentis* in an HGT event. The lack of similarity between the DHC BGC of *A. terreus* and any other in the could indicate a case of convergent evolution, where *P. tritici-repentis* and *A. terreus* acquired the ability to produce DHC completely separate from one another. Alternatively, the results could indicate that the *P. tritici-repentis, A. cinerariae, A. alternata, A. arborescens,* and *A. terreus* all share a distant common ancestor that has lost the ability to produce the curvularins over time. As the effects of the curvularins warrant its classification as a VF, this would completely oppose the traditional view of evolution by positive selection.

4.6 References

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H., et al. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 49, W29–W35. doi:10.1093/nar/gkab335.

Caputo, O., and Viola, F. (1977). Isolation of a, beta-dehydrocurvularin from Aspergillus aureofulgens. *Planta Med.* 31, 31–32. doi:10.1055/S-0028-1097485.

Chooi, Y.-H., and Tang, Y. (2012). Navigating the fungal polyketide chemical space: from genes to molecules. *J. Org. Chem.* 77, 9933–9953. doi:10.1021/jo301592k.

Cochrane, R. V. K., Gao, Z., Lambkin, G. R., Xu, W., Winter, J. M., Marcus, S. L., et al. (2015). Comparison of 10,11-Dehydrocurvularin Polyketide Synthases from *Alternaria cinerariae* and Aspergillus terreus Highlights Key Structural Motifs. *Chembiochem* 16, 2479–2483. doi:10.1002/cbic.201500428.

Crawford, J. M., Korman, T. P., Labonte, J. W., Vagstad, A. L., Hill, E. A., Kamari-Bidkorpeh, O., et al. (2009). Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature* 461, 1139–1143. doi:10.1038/nature08475.

Crawford, J. M., Thomas, P. M., Scheerer, J. R., Vagstad, A. L., Kelleher, N. L., and Townsend, C. A. (2008). Deconstruction of iterative multidomain polyketide synthase function. *Science* 320, 243–246. doi:10.1126/science.1154711.

Crawford, J. M., and Townsend, C. A. (2010). New insights into the formation of fungal aromatic polyketides. *Nat. Rev. Microbiol.* 8, 879–889. doi:10.1038/nrmicro2465.

Foulke-Abel, J., and Townsend, C. A. (2012). Demonstration of starter unit interprotein transfer from a fatty acid synthase to a multidomain, nonreducing polyketide synthase. *Chembiochem* 13, 1880–1884. doi:10.1002/cbic.201200267.

Gerlt, J. A. (2017). Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence–Function Space and Genome Context to Discover Novel Functions. *Biochemistry* 56, 4293–4308. doi:10.1021/acs.biochem.7b00614.

Gerlt, J. A., Bouvier, J. T., Davidson, D. B., Imker, H. J., Sadkhin, B., Slater, D. R., et al. (2015). Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta - Proteins Proteomics* 1854, 1019–1037. doi:https://doi.org/10.1016/j.bbapap.2015.04.015.

Ghisalberti, E. L., and Rowland, C. Y. (1993). 6-chlorodehydrocurvularin, a new metabolite from cochliobolus spicifer. *J. Nat. Prod.* 56, 2175–2177. doi:10.1021/NP50102A022.

Gutiérrez, M., Theoduloz, C., Rodríguez, J., Lolas, M., and Schmeda-Hirschmann, G. (2005). Bioactive Metabolites from the Fungus Nectria galligena, the Main Apple Canker Agent in Chile. *J. Agric. Food Chem.* 53, 7701–7708. doi:10.1021/jf051021l.

He, J., Wijeratne, E. M. K., Bashyal, B. P., Zhan, J., Seliga, C. J., Liu, M. X., et al. (2004). Cytotoxic and Other Metabolites of Aspergillus Inhabiting the Rhizosphere of Sonoran Desert Plants. *J. Nat. Prod.* 67, 1985–1991. doi:10.1021/np040139d.

Horgusluoglu, E, Neff, R, Song, W-M, et al. (2021). Integrative metabolomics-genomics approach reveals key metabolic pathways and regulators of Alzheimer's disease. *Alzheimer's Dement*. 1- 19. https://doi.org/10.1002/alz.12468

Hyeon, S.-B., Ozaki, A., Suzuki, A., and Tamura, S. (1976). Isolation of αβ-Dehydrocurvularin and β-Hydroxycurvularin from *Alternaria* tomato as Sporulation-suppressing Factors. *Agric. Biol. Chem.* 40, 1663–1664. doi:10.1080/00021369.1976.10862283.

Jeon, Y.-T., Ryu, K.-H., Kang, M.-K., Park, S.-H., Yun, H., QT, P., et al. (2010). Alternariol monomethyl ether and α, β-dehydrocurvularin from endophytic fungi *Alternaria* spp. Inhibit

appressorium formation of Magnaporthe grisea. *J. Korean Soc. Appl. Biol. Chem.* 53, 39–42. doi:10.3839/JKSABC.2010.007.

Kusano, M., Nakagami, K., Fujioka, S., Kawano, T., Shimada, A., and Kimura, Y. (2003). Pratylenchus penetrans from the Fungus Aspergillus sp. *Biosci. Biotechnol. Biochem* 67, 1413–1416. Available at: https://academic.oup.com/bbb/article/67/6/1413/5944393 [Accessed August 11, 2021].

Li, Y., Image, I. I., Xu, W., Image, I., Tang, Y., and Image, I. (2010). Classification, prediction, and verification of the regioselectivity of fungal polyketide synthase product template domains. *J. Biol. Chem.* 285, 22764–22773. doi:10.1074/jbc.M110.128504.

Moolhuijzen, P., See, P. T., Hane, J. K., Shi, G., Liu, Z., Oliver, R. P., and Moffat, C.S. (2018). Comparative genomics of the wheat fungal pathogen Pyrenophora tritici-repentis reveals chromosomal variations and genome plasticity. *BMC Genomics* 19, 1–23. doi:10.1186/S12864-018-4680-3.

Musgrave, O. C. (1956). 828. Curvularin. Part I. Isolation and partial characterisation of a metabolic product from a new species of Curvularia. *J. Chem. Soc.*, 4301–4305. doi:10.1039/JR9560004301.

Reyes, A., Alves, J. M. P., Durham, A. M., and Gruber, A. (2017). Use of profile hidden Markov models in viral discovery: current insights. *Adv. Genomics Genet.* 7, 29–45. doi:10.2147/AGG.S136574.

Robeson, D. J., Strobel, G. A., and Strange, R. N. (1985). The identification of a major phytotoxic component from *Alternaria macrospora* as αβ-dehydrocurvularin. *J. Nat. Prod.* 48, 139–141. Available at: https://pubs.acs.org/sharingguidelines [Accessed August 10, 2021].

Santagata, S., Xu, Y.-M., Wijeratne, E. M. K., Kontnik, R., Rooney, C., Perley, C. C., et al. (2012). Using the heat-shock response to discover anticancer compounds that target protein homeostasis. *ACS Chem. Biol.* 7, 340–349. doi:10.1021/cb200353m.

Vurro, M., Evidente, A., Andolfi, A., Chiara Zonno, M., Giordano, F., and Motta, A. (1998). Brefeldin A and α,β-dehydrocurvularin, two phytotoxins from *Alternaria zinniae*, a biocontrol agent of Xanthium occidentale. *Plant Sci.* 138, 67–79. doi:https://doi.org/10.1016/S0168-9452(98)00131-9.

Wang, B., Guo, F., Huang, C., and Zhao, H. (2020). Unraveling the iterative type I polyketide synthases hidden in Streptomyces. *Proc. Natl. Acad. Sci.* 117, 8449–8454. doi:10.1073/PNAS.1917664117.

Wang, M., Zhou, H., Wirz, M., Tang, Y., and Boddy, C. N. (2009). A thioesterase from an iterative fungal polyketide synthase shows macrocyclization and cross coupling activity and may play a role in controlling iterative cycling through product offloading. *Biochemistry* 48, 6288–6290. doi:10.1021/bi9009049.

Wiese, J., Ohlendorf, B., Blümel, M., Schmaljohann, R., and Imhoff, J. F. (2011). Phylogenetic Identification of Fungi Isolated from the Marine Sponge Tethya aurantium and Identification of Their Secondary Metabolites. *Mar. Drugs* 9. doi:10.3390/md9040561.

Witte, T., Harris, L., Nguyen, H., Hermans, A., Johnston, A., Sproule, A., et al. (2020). Apicidin biosynthesis is linked to accessory chromosomes in *Fusarium Poae* isolates. *BMC Genomics* Preprint. doi:10.21203/rs.3.rs-116075/v1.

Witte, T. E., Villeneuve, N., Boddy, C. N., and Overy, D. P. (2021). Accessory Chromosome-Acquired Secondary Metabolism in Plant Pathogenic Fungi: The Evolution of Biotrophs Into Host-Specific Pathogens . *Front. Microbiol.* 12, 700. Available at:

https://www.frontiersin.org/article/10.3389/fmicb.2021.664276.

Xiang, C., Liu, Y., Liu, S.-M., Huang, Y.-F., Kong, L.-A., Peng, H., et al. (2020). αβ-Dehydrocurvularin isolated from the fungus *Aspergillus welwitschiae* effectively inhibited the behaviour and development of the root-knot nematode Meloidogyne graminicola in rice roots. *BMC Microbiol. 2020 201* 20, 1–10. doi:10.1186/S12866-020-01738-2.

Xu, Y., Espinosa-Artiles, P., Schubert, V., Xu, Y., Zhang, W., Lin, M., et al. (2013). Characterization of the biosynthetic genes for 10,11-dehydrocurvularin, a heat shock response-modulating anticancer fungal polyketide from Aspergillus terreus. *Appl. Environ. Microbiol.* 79, 2038–2047. doi:10.1128/AEM.03334-12.

Yoon, B.-J. (2009). Hidden Markov Models and their Applications in Biological Sequence Analysis. *Curr. Genomics* 10, 402–415. doi:10.2174/138920209789177575.

Zallot, R., Oberg, N., and Gerlt, J. A. (2019). The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* 58, 4169–4182. doi:10.1021/acs.biochem.9b00735.

Zallot, R., Oberg, N. O., and Gerlt, J. A. (2018). 'Democratized' genomic enzymology web tools for functional assignment. *Curr. Opin. Chem. Biol.* 47, 77–85. doi:https://doi.org/10.1016/j.cbpa.2018.09.009.

Zhan, J., Wijeratne, E. M. K., Seliga, C. J., Zhang, J., Pierson, E. E., PIERSON III, L. S., et al. (2004). A New Anthraquinone and Cytotoxic Curvularins of a Penicillium sp. from the Rhizosphere of Fallugia paradoxa of the Sonoran Desert. *J. Antibiot. (Tokyo).* 57, 341–344. doi:10.7164/ANTIBIOTICS.57.341.

4.7 Appendix C

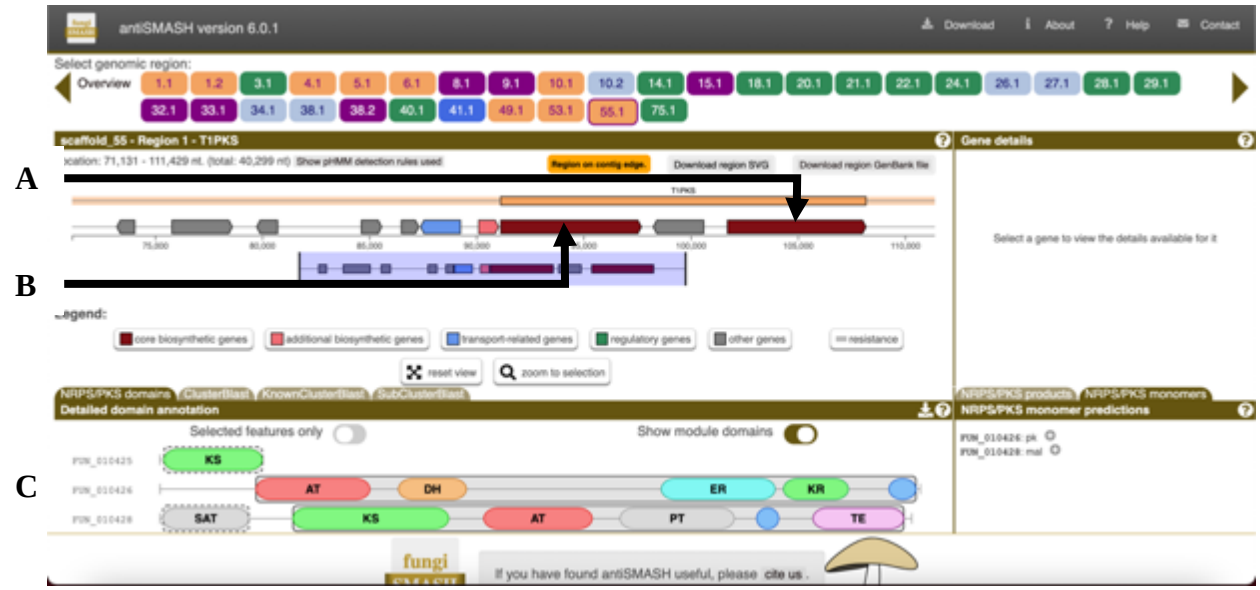

Figure 4.8. Screenshot of the antiSMASH v6.0.1 user interface, whereby the T1PKS region that denotes the DHC BGC is being highlighted, in a utilized *A. arborescens* isolate (KAS5321). The non-reducing PKS (nrPKS) and highly-reducing PKS (hrPKS) of the BGC are denoted by **A**, and **B,** respectively**.** The domains that comprise the genes of the gene cluster (the nrPKS and hrPKS) are denoted by **C** (with the nrPKS being FUN_010428 and the hrPKS being FUN_010426).

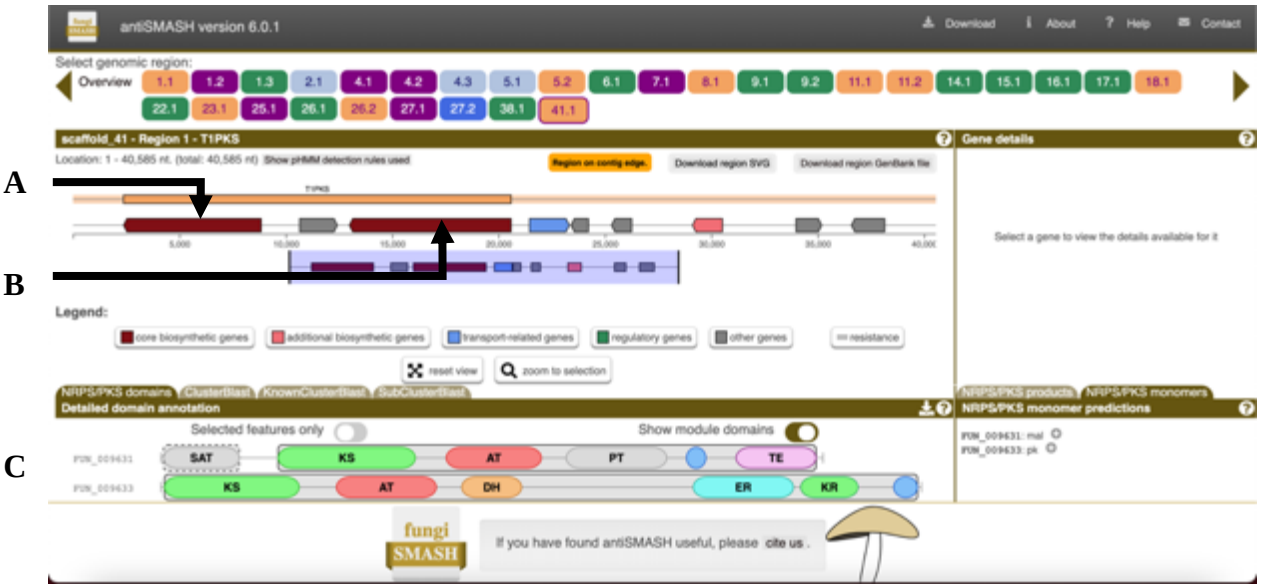

Figure 4.9. Screenshot of the antiSMASH v6.0.1 user interface, whereby the T1PKS region that denotes the DHC BGC is being highlighted, in a utilized *A. alternata* isolate (KAS5743). The non-reducing PKS (nrPKS) and highly-reducing PKS (hrPKS) of the BGC are denoted by **A**, and **B,** respectively**.** The domains that comprise the genes of the gene cluster (the

nrPKS and hrPKS) are denoted by **C** (with the nrPKS being FUN_009633 and the hrPKS being FUN_009631).

Table 4.1. The sequences of highest similarity to the non-reducing polyketide synthase (nrPKS) of the dehydrocurvularin (DHC) biosynthetic gene cluster (BGC) of the *A. arborescens* isolate, from which, DHC was obtained (DET2035). Sequence similarity metrics were query coverage (%), E-value, and percent identity (%), as according to BLASTn.

Description	Scientific Name	Query Coverage	E-value	Percent Identity	Accession
Alternaria cinerariae strain ATCC 11784 non-reducing polyketide synthase (Dhc5) mRNA, complete cds	*Alternaria cinerariae*	99%	0	94.90%	KT271474.1
Alternaria solani isolate NL03003 chromosome 8, complete sequence	*Alternaria solani*	99%	0	94.67%	CP022031.1
Pyrenophora tritici-repentis Pt-1C-BFP conidial yellow pigment biosynthesis polyketide synthase, mRNA	*Pyrenophora tritici-repentis Pt-1C-BFP*	94%	0	88.49%	XM_001934683.1
Aspergillus terreus strain AH-02-30-F7 dehydrocurvularin biosynthetic gene cluster, complete sequence	*Aspergillus terreus*	25%	0	77.32%	JX971534.1
Aspergillus felis strain FM324 chromosome 7	*Aspergillus felis*	14%	1.00E-155	78.04%	CP066509.1

Table 4.2. The sequences of highest similarity to the highly-reducing polyketide synthase (hrPKS) of the dehydrocurvularin (DHC) biosynthetic gene cluster (BGC) of the *Alternaria arborescens* isolate, from which, DHC was obtained (DET2035). Sequence similarity metrics were query coverage (%), E-value, and percent identity (%), as according to BLASTn.

Description	Scientific Name	Query Coverage	E-value	Percent Identity	Accession
Alternaria cinerariae strain ATCC 11784 highly reducing polyketide synthase (Dhc3) mRNA, complete cds	*Alternaria cinerariae*	100%	0	95.69%	KT271472.1
Pyrenophora tritici-repentis Pt-1C-BFP phenolpthiocerol synthesis polyketide synthase ppsB, mRNA	*Pyrenophora tritici-repentis Pt-1C-BFP*	99%	0	88.06%	XM_001934685.1
Alternaria solani isolate NL03003 chromosome 8, complete sequence	*Alternaria solani*	99%	0	95.50%	CP022031.1
Aspergillus novofumigatus IBT 16806 phenolpthiocerol synthesis polyketide synthase ppsB (P174DRAFT_515935), partial mRNA	*Aspergillus novofumigatus IBT 16806*	34%	0	74.60%	XM_024832249.1
Aspergillus niger CBS 513.88 polyketide synthase, mRNA	*Aspergillus niger CBS 513.88*	0%	1.00E-06	95.45%	XM_001388518.2

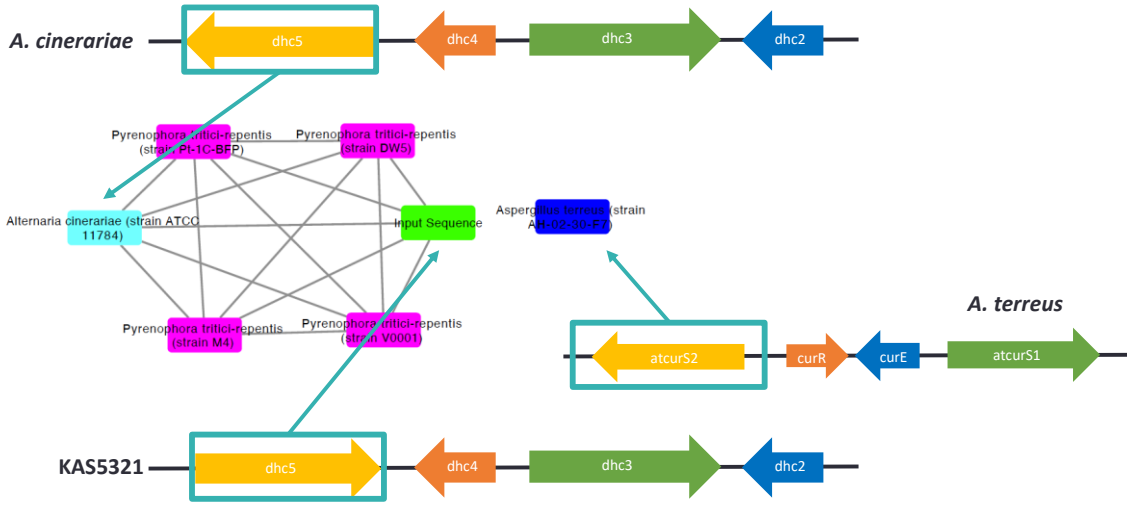

Figure 4.10. The sequence similarity network generated from the use of the nrPKS of the DHC BGC of KAS5321 as the query sequence, with 10 nodes represented. Each network (or cluster of nodes) represents the nodes that could be connected with highest similarity to one another (with > 85% identity and >= 1150 alignment score) (except the singleton node of *A. terreus* (L7XAV2)). Nodes highlighted in bright blue denote those from *Alternaria*, pink denotes those from *Pyrenophora*, dark blue denotes those from *Aspergillus*. The input sequence is denoted by the bright green node.

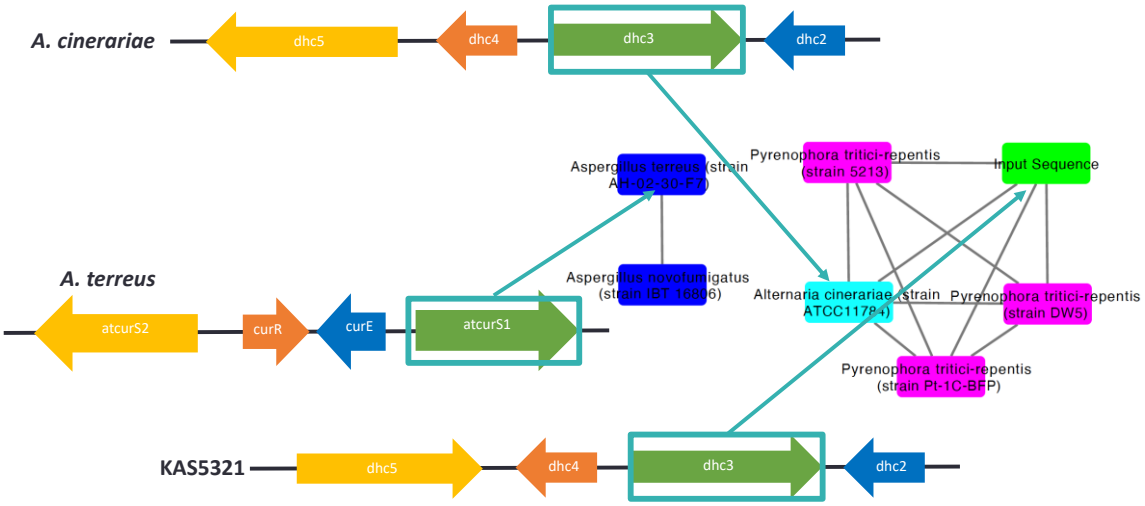

Figure 4.11. The sequence similarity network generated from the use of the hrPKS of the DHC BGC of KAS5321 as the query sequence, with 7 nodes represented. Each network (or cluster of nodes) represents the nodes that could be connected with highest similarity to one another (with > 85% identity and >= 1150 alignment score). Nodes highlighted in bright blue denote those from *Alternaria*, pink denotes those from *Pyrenophora*, dark blue denotes those from *Aspergillus*, and the bright green node represents the input sequence.

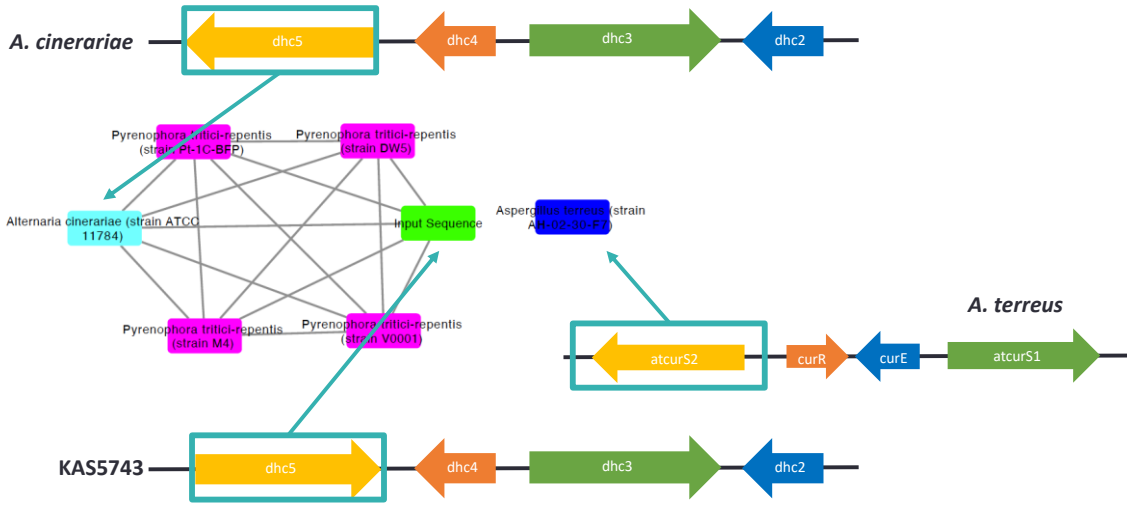

Figure 4.12. The sequence similarity network generated from the use of the nrPKS of the DHC BGC of KAS5743 as the query sequence, with 10 nodes represented. Each network (or cluster of nodes) represents the nodes that could be connected with highest similarity to one another (with > 85% identity and >= 1150 alignment score) (except the singleton node of *A. terreus* (L7XAV2)). Nodes highlighted in bright blue denote those from *Alternaria*, pink denotes those from *Pyrenophora*, dark blue denotes those from *Aspergillus*. The input sequence is denoted by the bright green node.

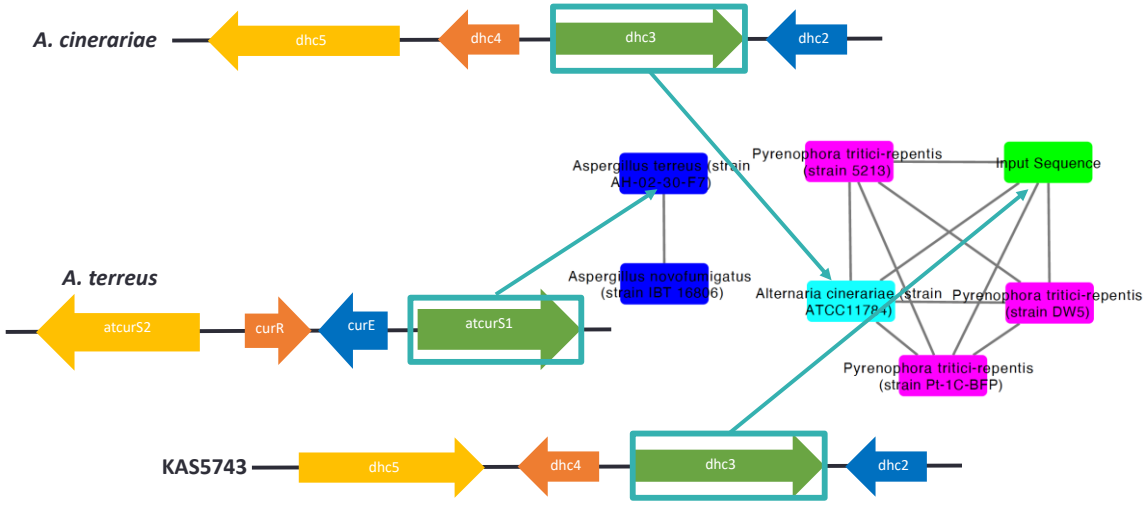

Figure 4.13. The sequence similarity network generated from the use of the hrPKS of the DHC BGC of KAS5743 as the query sequence, with 7 nodes represented. Each network (or cluster of nodes) represents the nodes that could be connected with highest similarity to one another (with > 85% identity and >= 1150 alignment score). Nodes highlighted in bright blue denote those from *Alternaria*, pink denotes those from *Pyrenophora*, dark blue denotes those from *Aspergillus*. The input sequence is denoted by the bright green node.

Chapter 5

5 Summary and Future Directions

The overall objective of the present thesis was to use untargeted LC-MS metabolomics to study the metabolome of 60 stains of uncharacterized *Alternaria spp.*, selected from various species-groups in the *Alternaria* genus. In turn, metabolomics was utilized to understand trends in *Alternaria* SM expression. Specifically, 39 strains were selected on the basis of belonging to either one of two species-groups that are not thought to be resolvable form one another.

Towards accomplishing this task, a massive dataset was produced which, when mined, revealed the presence of several unique mass features. Subsequent targeted metabolomic work was performed to determine that it was 5 members of the curvularins family that displayed unique production patterns. The work performed for this thesis focused on this suite of curvularins, but many more mass features exhibited unique production patterns. Comparison of DET2035, KAS5321, and KAS5743 to the rest of the *Alternata* group revealed 182 mass features that exhibited a statistically significant difference in production, of which, only about 15% could be linked to the curvularins. Of those 182 features, similar fragmentation patterns as were utilized to annotate the curvularins, were detected in many other features. As such, future purification work could result in the discovery of several novel curvularins or an increased understanding of their synthesis.

Historically, metabolomic studies performed on *Alternaria* have focused on assessing the validity of applying a chemotaxonomic approach to *Alternaria* classification, assessing the

presence of emerging mycotoxins in fungal and food/feed samples, or biomarker discovery. The work performed in the present thesis serves to validate looking at multiple species to identify SMs produced by only a subset as a strategy for finding interesting functional SMs. It has set the foundation for future work of elucidating the structure and the identification of those remaining features, along with the hundreds of others exhibiting unique detection patterns. This could help to better understand the pathogenic potential that *Alternaria* pose to plant and animal life.

To build on the findings of **Chapter 2**, a scale-up growth fermentation of DET2035 was performed with the goal of determining the identity of the metabolites that distinguish DET2035, KAS5321, and KAS5743 as unique from the other *Alternaria spp*. studied. The results obtained from processing the bulk extract served to confirm that the detected molecular "family" was the curvularins. This could be confidently concluded based on that a pure sample of DHC was obtained (proven by NMR results matching those reported in literature) (Hassan 2007; Aly et al., 2010). Additionally, the *m/z* of the mass features matched those of DHC-analogues (within 3 ppm) and the MS/MS fragments of the mass features served to suggest they are all of the same family (as per that they all share many of the same fragments). This is a significant finding as it is the first report of DHC and any DHC-analogues being isolated from an *A. arborescens* isolate. Several solvent systems have been utilized to isolate the curvularins in the past. The experimentation carried out for **Chapter 3** of the present thesis denotes the first report of using 5% MeOH in Chloroform to isolate a pure sample of DHC.

Additionally, this is the first report of sumalarin C in an *Alternaria* species. Sumalarin C has, to date, only been reported in an *Aspergillus sumatrense* isolate, in a study where they only report

its characterization (Meng et al., 2013). Its biosynthesis, therefore, has not been explicitly detailed. Future work should be to elucidate the biosynthetic pathway that is responsible for the production of sumalarin C. Sumalarin C has been reported to exhibit potent cytotoxicity against some of tested tumor cell lines, though the its mode of action is not known. This is much the same case for the rest of the curvularins (Meng et al., 2013). Many have been reported to exhibit a whole range of toxic effects, but the mechanism by which they do so remains largely unknown. As such, future work should be towards determining the bioactivity and mode of action of the curvularins. Future bioassays could be conducted to gain a better understanding of the bioactivity of the curvularins in plants utilizing model plant organisms like *Arabadopsis thaliana*, or perhaps the triticale plant to mimic natural conditions for the utilized *Alternaria* isolates. To gain a better understanding of the bioactivity of the curvularins in animals, numerous in vitro or in vivo assays could be utilized.

To date, the DHC BGC has only been well characterized in *A. cinerariae*, and *A. terreus*. Additionally, the DHC BGC has been found in *P. tritici-repentis*, despite *Pyrenophora* having never been reported to produce curvularins. The findings of **Chapter 4** therefore represent the first reports of the DHC BGC in an *A. arborescens* and non-pathotype *A. alternata* isolate. Additionally, this study represents the first reported application of a sequence similarity network to the DHC BGC. In doing so, an interesting relatedness of the DHC BGC was implicated between the three studied *Alternaria* isolates and various *P. tritici-repentis* isolates. This may serve to suggest that the origin of the DHC BGC in *Alternaria* could be traced back to a HGT event that occurred when a *P. tritici-repentis* and *A. arborescens*, *A. alternata*, or *A. cinerariae* were competing for a triticale plant (**Figure *5.1***). Alternatively, it could indicate that they all

share a distant common ancestor that has lost the ability to produce the curvularins in all but a small sub-population. As the curvularins are believed to be VFs, this opposes what would be expected – that the BGC be actively upheld by positive selection pressure. Further work should therefore be conducted to elucidate the relatedness and origin of the BGC in both genera. Future work should also be to upload the genomes of the three sequenced *Alternaria spp*. to NCBI, in turn, doubling the amount of publicly-available genomes containing the DHC BGC.

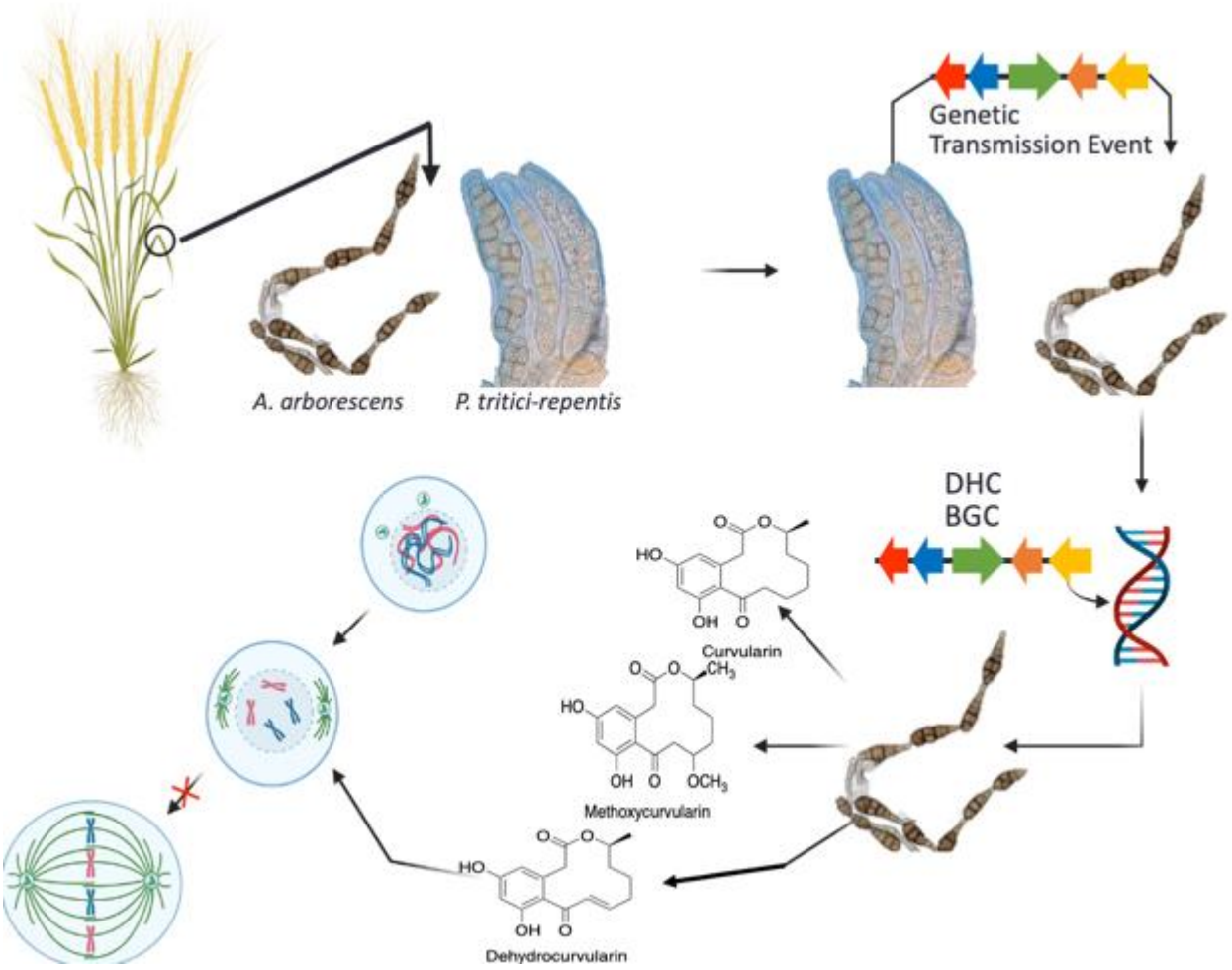

Figure 5.1. An illustration depicting the introduction of the DHC BGC into the genome of a pathogenic *Alternaria* isolated from a triticale plant.

5.1 References

Aly, A. H., Debab, A., Edrada-Ebel, R. A., Müller, W. E. G., Kubbutat, M. H. ., Wray, V., et al. (2010). Protein kinase inhibitors and other cytotoxic metabolites from the fungal endophyte Stemphylium botryosum isolated from Chenopodium album. *Mycosphere* 1, 153–162.

Hassan, A. E. H. . (2007). Novel natural products from endophytic fungi of Egyptian medicinal plants : chemical and biological characterization. Available at: http://www.secheresse.info/spip.php?article5992 [Accessed August 28, 2021].

Meng, L.-H., Li, X.-M., Lv, C.-T., Li, C.-S., Xu, G.-M., Huang, C.-G., et al. (2013). Sulfur-Containing Cytotoxic Curvularin Macrolides from Penicillium sumatrense MA-92, a Fungus Obtained from the Rhizosphere of the Mangrove Lumnitzera racemosa. *J. Nat. Prod*. 76, 2145–2149. doi:10.1021/NP400614F.

6 Supplementary Information

6.1.1 Fungal Strains

Table 6.1. A list of the strains of the utilized for the experimentation conducted for the present chapter, with their strain ID and the species section to which they can be taxonomically organized (based on phylogenetic work conducted by the laboratories of Dr. J. Dettman)

Strain	Species section
KAS1274	*sect. Alternaria A. alternata*
DET2071	*sect. Alternaria A. alternata*
DET2076	*sect. Alternaria A. alternata*
DET2080	*sect. Alternaria A. alternata*
DET2092	*sect. Alternaria A. alternata*
DET2110	*sect. Alternaria A. alternata*
DET2123	*sect. Alternaria A. alternata*
KAS5299	*sect. Alternaria A. alternata*
KAS5303	*sect. Alternaria A. alternata*
KAS5306	*sect. Alternaria A. alternata*
KAS5313	*sect. Alternaria A. alternata*
KAS5321	*sect. Alternaria A. alternata*
KAS5372	*sect. Alternaria A. alternata*
KAS5386	*sect. Alternaria A. alternata*
KAS5394	*sect. Alternaria A. alternata*
KAS5428	*sect. Alternaria A. alternata*
KAS5456	*sect. Alternaria A. alternata*
KAS5468	*sect. Alternaria A. alternata*
KAS5485	*sect. Alternaria A. alternata*
KAS5489	*sect. Alternaria A. alternata*
KAS5497	*sect. Alternaria A. alternata*
KAS5499	*sect. Alternaria A. alternata*
KAS5513	*sect. Alternaria A. alternata*
KAS5515	*sect. Alternaria A. alternata*
KAS5516	*sect. Alternaria A. alternata*
KAS6096	*sect. Alternaria A. alternata*
KAS6097	*sect. Alternaria A. alternata*
DET2042	*sect. Alternaria infectoriae*
DET2072	*sect. Alternaria infectoriae*
DET2103	*sect. Alternaria infectoriae*

DET2104	*sect. Alternaria infectoriae*
DET2105	*sect. Alternaria infectoriae*
DET2106	*sect. Alternaria infectoriae*
DET2107	*sect. Alternaria infectoriae*
KAS5446	*sect. Alternaria infectoriae*
KAS5449	*sect. Alternaria infectoriae*
KAS5470	*sect. Alternaria infectoriae*
KAS5474	*sect. Alternaria infectoriae*
KAS5477	*sect. Alternaria infectoriae*
KAS5486	*sect. Alternaria infectoriae*
KAS5488	*sect. Alternaria infectoriae*
KAS5506	*sect. Alternaria infectoriae*
KAS5298	*sect. Alternaria Ulocladioides*
KAS5364	*Pithomyces*
KAS5756	*sect. Alternaria Ulocladioides*
KAS5799	*sect. Alternaria Ulocladium*
KAS5825	*sect. Alternaria Pseudoulocladium*
KAS5913	*sect. Alternaria Ulocladioides*
DET2035	*sect . Alternaria A. arborescens*
DET2078	*sect . Alternaria A. arborescens*
DET2085	*sect . Alternaria A. arborescens*
DET2119	*sect . Alternaria A. arborescens*
KAS5275	*sect . Alternaria A. arborescens*
KAS5321	*sect . Alternaria A. arborescens*
KAS5368	*sect . Alternaria A. arborescens*
KAS5398	*sect . Alternaria A. arborescens*
KAS5399	*sect . Alternaria A. arborescens*
KAS5521	*sect . Alternaria A. arborescens*
KAS5743	*sect . Alternaria A. arborescens*
KAS5762	*sect . Alternaria A. arborescens*

Table 6.2. The test statistic, p-value (p), logarithmic p value ($-Log_{10}(p)$), and false discovery rate of the mass features highlighted in Figure 2.13.

	Test statistic	P-value (p)	-Log10(p)	False Discovery Rate
RT3.69_195.02872	-35.032	2.20E-76	75.659	1.81E-74
RT3.69_328.10654	-93.615	2.60E-140	139.59	3.42E-137
RT3.7_309.13205	-33.094	6.04E-73	72.219	4.19E-71
RT3.71_273.11148	-38.643	2.04E-82	81.691	4.38E-80
RT3.87_195.02898	-35.907	6.87E-78	77.163	6.95E-76
RT3.89_273.11169	-36.398	1.01E-78	77.797	1.21E-76
RT3.90_259.05958	-2.7638	0.006393	2.1943	0.046482

RT3.90_169.04926	-37.307	3.07E-80	79.513	4.04E-78
RT3.91_413.12519	-38.607	2.33E-82	81.633	4.38E-80
RT3.91_225.11174	-7.3407	1.06E-11	10.977	1.34E-10
RT3.92_123.08033	-36.154	2.61E-78	77.583	2.86E-76
RT3.92_245.08104	-35.228	1.00E-76	75.998	8.80E-75
RT3.94_435.10862	-60.179	7.71E-111	110.11	3.38E-108
RT4.14_307.18974	-4.1789	4.84E-5	4.3156	0.000461
RT4.15_345.13078	-12.596	1.25E-25	24.903	2.26E-24
RT4.37_195.02882	-21.028	1.19E-47	46.923	4.24E-46
RT4.37_249.11185	-11.692	3.79E-23	22.421	6.57E-22
RT4.38_273.11076	-22.286	1.23E-50	49.911	5.21E-49
RT4.38_123.04361	-19.332	1.75E-43	42.758	5.10E-42
RT4.38_169.0493	-21.878	1.12E-49	48.95	4.61E-48
RT4.38_193.0493	-20.527	1.96E-46	45.708	6.29E-45
RT4.38_205.04916	-21.316	2.43E-48	47.615	9.40E-47
RT4.38_245.11696	-22.747	1.04E-51	50.985	4.54E-50
RT4.38_291.12239	-19.914	6.23E-45	44.205	1.95E-43
RT4.47_293.13755	-9.2604	1.41E-16	15.582	2.13E-15
RT4.48_275.12708	-9.1447	2.84E-16	15.547	4.20E-15

6.1.2 R Coding

Any text surrounded by ## at both ends, denotes text embedded in the code (comments). The absence of "##" means that the code was executed to yield the result specified in the comments.

```
## First step is to create the mSet Object, specifying that the data to be uploaded ##
## is a peak table ("pktable") and that statistical analysis will be performed ("stat"). ##
mSet <- InitDataObjects(data.type="pktable", anal.type="stat", paired=FALSE)

## Second step is to import the processed peak data matrix ##
## For this script, will be using my Alternata group peak list, containing 2 classes ##
```

```
## class 1 == the samples that did not produce 291.122, class 2 == the samples that did
(KAS3521, KA5743, DET2035) ##
mSet<-Read.TextData(mSetObj=mSet,
filepath="/Users/Nicolas/Desktop/WorkingDir/Alternata_Non_Binary_CYS80_2Classes.csv",
format="colu", lbl.type="disc")

## The third step is to start Data processing with MetaboAnalystR. To do so, must ##
## first do a SanityCheck ##
mSet<-SanityCheckData(mSet)

## If it passes the SanityCheck, can now perform data filtering, scaling and normalization ##
## As the goal would be to normalize my peak list relative to an internal standard (in the ref
parameter), ##
## had to apply the PreparePrenormdata function, and then the normalization function was
applied ##
mSet <- PreparePrenormData(mSet)
mSet <- Normalization(mSetObj=mSet, rowNorm='CompNorm', transNorm='NULL',
scaleNorm='NULL', ref='RT2.85_195.08721', ratio=FALSE, ratioNum=0)

## MULTIVARIATE ANALYSES ##

## Once the data had been normalized, net step was to preform multivariate statistical analyses
on the data ##
```

```
# To do so, first used the PCA method built into the MetaboAnalystR package which ##
## performs PCA analysis, obtains variance explained, stores that all as a PCA object ##
mSet <- PCA.Anal(mSet)


## To view the first 5 pairwise PC plots, can use the following code##
mSet <- PlotPCAPairSummary(mSetObj=mSet, imgName='pca_pair_vis', format='png', dpi=72,
width=NA,pc.num=5)

## To then view the scores plots, comparing PC1 against PC2, can use the following code. ##
mSet <- PlotPCA2DScore(mSetObj=mSet, imgName='pca_scores_plots', format='png', dpi=72,
width=NA, pcx=1, pcy=2, reg=0.95, show=0,grey.scale=0)

## In order to display the sample names in the scores plots, comparing PC1 against PC2, can use
the following code ##
## notice the only difference is to change the show variable to a 1 ##
mSet <- PlotPCA2DScore(mSetObj=mSet, imgName='pca_scores_plots', format='png', dpi=72,
width=NA, pcx=1, pcy=2, reg=0.95,show=1, grey.scale=0)

## To then view the Loadings plot, comparing Loadings1 (x-axis) to Loading2 (y-axis) ##
mSet <- PlotPCALoading(mSetObj=mSet, imgName='pca_loadings_plots', format='png',
dpi=72, width=NA, inx1=1, inx2=2)
```

Finally, to view the biplot, where PC1 and PC2 are plotted against Loadings1 and Loadings2, can use the following code

mSet <- PlotPCABiplot(mSetObj=mSet, imgName='Biplots', format='png', dpi=72, width=NA, inx1=1, inx2=2)

UNIVARIATE ANALYSES

In order to apply T-Test analysis on the processed peak data, used the following code

mSet <- Ttests.Anal(mSetObj=mSet, nonpar=F, threshp = 0.05, paired=FALSE, equal.var=TRUE, all_results = FALSE)

to then visualize the T-test analysis, used the following code

mSet <- PlotTT(mSetObj=mSet, imgName='T-Tests', format='png', ddpi=72, width=NA)

to then visualize the production patterns of certain mass features, used the PlotCmpdView function (as follows)

to view other mass features then that which is exemplified below, you simply need change the cmpdNm parameter to the mass feature of interest

mSet <- PlotCmpdView(mSetObj=mSet, cmpdNm='RT4.38_123.04361', format='png', dpi=72, width=NA)

In order to assess if a dichotamy was present in the dataset, the dataset

was converted to binary (in the fashion explained in the methods of this paper).

```
## After importing the binary dataset into the R environment, the goal was to ##
## create a scores plot and heatmap. To accomplish these tasks the following ##
## code was used ##

## The first step was to create the nSet object, specifying the value to be uploaded ##
nSet <- InitializeDataObjects("pktable", "stat", FALSE)

# The next step was to then import the peak data matrix #
 nSet<-Read.TextData(mSetObj=nSet,
filepath="/Users/Nicolas/Desktop/WorkingDir/ConsensusMatrixForMA.csv", format="colu",
lbl.type="disc")

## The third step was, again, to do a SanityCheck ##
nSet<-SanityCheckData(nSet)

## As the data was binary, there was no need to scale, normalize or filter the data ##
## Next step was therefore to go right into PCA and generating scores plots ##
nSet <- PCA.Anal(nSet)
nSet <- PlotPCA2DScore(mSetObj=nSet, imgName='consens_scores_plots', format='png',
dpi=72, width=NA, pcx=1, pcy=2, reg=0.95,show=1, grey.scale=0)

## Finally, to generate a heatmap, the following code was utilized ##
```

```
nSet <- PlotHeatMap(mSetObj = nSet, imgName="consens_heatmap", format="png", dpi=72,
width=NA, dataOpt="raw", scaleOpt="none", smplDist="euclidean", clstDist="ward.D",
palette="gray", viewOpt="overview", rowV=T, colV=T, var.inx=NULL, border=T, grp.ave=F)
```