Polyketide Biosynthesis in Lichen Fungi *Cladonia uncialis*

by

MONA E ABDEL-HAMEED

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

Faculty of Science
University of Manitoba
Winnipeg, Manitoba, Canada

January 2015

Abstract

Lichens are known producers of a variety of secondary metabolites. Fungal polyketides constitute a large family of these secondary metabolites that have a high degree of structural diversity. Usnic acid is a lichen metabolite that has a broad range of biological activities. However the use of usnic acid in medical applications is limited due to the slow growth of the lichen in nature. In order to conduct in depth studies on compounds such as usnic acid it will be necessary to express their gene clusters in fast growing organisms. To achieve this goal, polyketide biosynthesis in the lichen fungi *Cladonia uncialis* has been investigated to identify the gene clusters that are responsible for secondary metabolites production.

This thesis reports the de novo whole-genome sequencing of the lichen *Cladonia uncialis*, in silico analysis of polyketide biosynthetic gene clusters, and putative identification and annotation of 56 different secondary metabolite gene clusters. The identified gene clusters include thirty two different type I polyketide synthase genes (non-reducing, partly reducing and highly reducing PKS genes) gene clusters besides two gene clusters of type III PKS gene, three independent, novel non-ribosomal peptide synthetases (NRPS) biosynthetic gene clusters, as well as seven noncanonical NRPS genes that did not contain condensation (C) domains, three polyketide non-ribosomal (PKS-NRP) hybrid gene clusters, one lanthipeptide synthase gene and six different terpene synthase gene clusters.

Out of 32 candidate genes a single PKS has been identified as being responsible for usnic acid biosynthesis.

The structure of one of the lichen non-reducing PKS genes that is responsible for production of a halogenated lichen metabolite was also studied. Based on the biosynthetic operations of the gene cluster as well as catalogued examples of halogenated polyketides isolated from lichen fungi to date, this study suggests that the gene cluster is a biosynthetic gene for an unidentified anthraquinone. The ketosynthase and the acyltransferase domains among two different non reducing PKS genes from *Cladonia uncialis* genome sequence are also studied. The amino acid sequences of the domains are confirmed by using mass spectrometry. Protein homology modeling was performed using the Swiss-Model server. The generated protein models were visualized using LASERGENE Protean 3D molecular visualization system.

Acknowledgments

First and foremost, I would like to thank my advisor Dr. John Sorensen, for giving me the opportunity to work on this project, and also for his support and confidence in my work. Dr. Sorensen patience and encouragement were invaluable to me throughout the course of this PhD thesis. He is such a great advisor to have. He guided me to perform to the best of my abilities and gave me opportunities and exposure I would never have, if I had not joined his group. I would also like to thank my committee member Dr. Michele D. Piercey-Normore for her support, guidance, patience; motivation and having her door all the time open for me. I would like to express my gratitude to my thesis Supervisory Committee Member Dr. Jorg Stetefeld for his support, motivation, help and assistance. I would like to thank my thesis Supervisory Committee Member Dr. Joe O'Neil for his good advice, valuable suggestions, ideas and enthusiasm.

I would like to thank Dr. Sean Mckenna's group for their help and assistance. Also I would like to thank Dr. Jennifer van Wijngaarden for giving me a space in her lab to finish writing my thesis. Special thanks to Dr. Lynda Donald for the MS analysis. Also Special thanks to Robert Bertrand, Miloslav Sailer, Lucile Jeusset for all their help, trouble-shooting ideas, lab assistance, enthusiasm and company. Thanks for all my lab mates in Dr. Sorensen's group and Dr. Piercey-Normore's group.

I would like to thank my family who has always been a source of encouragement throughout my life. Thanks to Abdelhalim and my children Zeyad, Mazen, Basim and Sarah for understanding the type of research that took me away from them. All thanks go to my loving mother for her sincere prayers for me and for my brother for his encouragment and support.

Contents

List of Tables

List of Figures

List of Symbols

µg	microgram
µL	microlitres
µm	micromoles
ACP	acyl carrier protein
antiSMASH	antibiotics & Secondary Metabolite Analysis SHell
AT	acyltransferase
C.uncialis	*Cladonia uncialis*
DEBS	6-Deoxyerythronolide B Synthase
Enz	enzyme
et al.	and others
FAS	fatty acid synthase
g	grams
H	hours
HPLC	high performance liquid chromatography
KS	ketosynthase
L	litter
mg	milligram

min	minutes
mL	millilitres
M-MLV	Moloney murine leukemia virus
MPAO	methylphloracetophenone oxidase
mpao	gene coding for MPAO
mpas	gene coding for MPAS
MPAS	methylphloracetophenone synthase
MS	mass spectrometry
NADPH	nicotinamide adenine dinucleotide phosphate
Ng	nanograms
nm	nanometers
NMR	nuclear magnetic resonance
NRPS	Non ribosomal peptides
PCR	polymerase chain reaction
PDA	potato dextrose agar
PDB	potato dextrose broth
pH	power of hydrogen
PKS	polyketide synthase
pmol	picomole
PT	product template
Rf	retention factor
RF	radio frequency
rpm	revolutions per minute

SAM	S-adeonsyl methionine
SAT	starter unit acyl-carrier protein transacylase
sec	second
sp.	Species
SUMO	Small Ubiquitin-like Modifier
TBE	tris (hydroxymethyl) aminomethane borate ethylenediaminetetraacetic acid
TE	thioesterase
TES	tris (hydroxymethyl) aminomethane ethylenediaminetetraacetic acid
TLC	thin layer chromatography
Tris-HCl	tris (hydroxymethyl) aminomethane hydrochloric acid
umol	micromoles
UV	ultraviolet
V	volts

List of Appendices

.

1 Chapter 1

1.1 Introduction

Natural products represent the main source of novel chemical compounds (Alvin A. *et al.,* 2014). These have played an important role in medicine since the 1940s (e.g. penicillin) and still serve as an important source for new medication (Newman *et al*. 2003). Despite the great synthetic diversity of medication that is derived from the development of high-throughput combinatorial chemistries, natural products are still considered to be extremely important for pharmaceutical companies. Natural products are likely to become even more important for the development of improved medicines to combat the adaptability of infectious bacteria. New and fast techniques of DNA sequencing and related genomics and bioinformatics tools are accelerating the discovery of new natural products and shortening drug discovery timelines (Alvin A. *et al.,* 2014, Newman D. J. *et al*., 2012, C. Walsh C T. 2008).

The research undertaken in this thesis was motivated by the goal to expand the natural product resources for pharmaceutical and industrial applications and to discover the gene cluster for one of the most important lichen metabolites, usnic acid. Usnic acid is one of the most intensively studied lichen secondary metabolites, its biological activity such as

antibacterial activity was recorded over 60 years ago (Shibata *et al*., 1948). Usnic acid also has activity against resistant strains of the *Mycobacterium tuberculosis* as well as antiproliferative, antiviral, anti-inflammatory and antiprotozoan activities (Ramos *et al*., 2010, De Carvalho *et al*., 2005, Vijayakumar *et al*., 2000).

For many years lichens have been used in traditional medicines and cosmetics (Agelet and Valles 2001, Prokasa *et al*. 1994), but today they are gaining interest in pharmaceutical research as a source of novel secondary metabolites. These lichen secondary metabolites are produced in the fungal partner of the lichen association, which utilizes the carbon fixed by the algal partner (Molnar *et al*., 2010). More than 1,050 different lichen secondary metabolites have been identified to date and about 18,500 lichen species are known (Molnar *et al*., 2010). Lichen metabolites exhibit manifold bioactivities including antibiotic, antimycobacterial, antiviral, antiinflammatory, analgesic, antipyretic, antiproliferative and cytotoxic effects (Muller 2001, Huneck 1999). The major problem of using lichens in a large scale is their slow growth and the difficulties in establishing pure cultures. According to Miao (2001), the main reason for lack of utilization of lichen resources is that lichens are slow-growing in nature, and so to harvest large amounts of lichen from nature is neither sustainable nor ecologically sensible. Although there have been successful attempts to grow lichen mycobionts in lab culture, the cultures often do not produce metabolites that resemble the compounds produced in the natural thalli which represent the vegetative part of the lichen (Stocker-Worgotter & Elix 2002, Huneck 1999, Culberson & Armaleo 1992, Crittenden & Porter 1991). Hence, transferring the lichen metabolite-producing genes into an easily grown heterologous host ap-

pears to be an attractive alternative approach (Muller 2001), as this might provide a sustainable source of useful lichen metabolites. Recently our understanding of the secondary metabolic pathways of various organisms as well as advances in molecular genetic techniques has opened up the possibility of accessing the lichen genomes directly to harness their potential for the production of useful secondary metabolites. Investigating the gene clusters for these secondary metabolites and cloning of metabolite-producing genes into heterologous hosts would not only enable expression of secondary metabolites that are normally found in lichen thalli but could also reveal silent metabolic pathways that are activated only under stress or extreme conditions (Gerke, J *et al*., 2012, Huneck 1999). Production of lichen metabolites in a fast-growing and well characterized heterologous host could also make fermentation scaling-up and titre improvement easier to achieve when potential therapeutic molecules are identified. Among the different secondary metabolic pathways in organisms, the genetics and biochemistry of polyketide biosynthesis are especially well-studied (Tabrez Khan S. *et al*., 2014). Prior knowledge, especially from other fungi, should be transferable to lichens, it is therefore reasonable to focus initial explorations on polyketide pathways in lichens. Furthermore, most unique lichen compounds, including usnic acid that has a wide range of biological activities are formed via the polyketide pathways (Da Costa Silva, J. *et al*., 2010). The tendency of polyketide synthase (PKS) genes, which code for the key enzymes involved in polyketide biosynthesis, to cluster with other genes within the same pathway (Keller & Hohn 1997) also eases the cloning and expression of all the biosynthetic genes required to produce the final products in the pathway. Despite the great benefits and commercial interests underlying

the investigation, the progress of research on the molecular biology of lichen secondary metabolism has been relatively slow compared with that in other free-living fungi. At the inception of this project in late 2007, there was only one complete lichen PKS gene sequence (Y.-H. Chooi *et al,* 2008) in the public domain, accession number (EF547512). The only other work was a study on the phylogeny of some lichen polyketide synthase genes inferred from the conserved β-ketoacyl synthase (KS) domain sequences (Grube & Blaha 2003). Understandably, progress was hampered due to the slow-growing nature of lichen and difficulty to culture the fungi. This project investigates and identifies the gene cluster of usnic acid which represents one of the most important lichen metabolites due to it is broad biological activities (Sokolov D. *et al*., 2012). Also an investigation was carried out on different polyketide gene clusters in lichenized fungi (*Cladonia uncialis*) in an attempt to overcome these difficulties. The ultimate aim is to use molecular genetic technologies to provide a sustainable source of lichen metabolites for natural product discovery and industrial applications. Next in this chapter is background information about the lichens, and the potential applications of the lichen metabolites. The function of polyketide synthases and their role in the biosynthesis of characteristic lichen compounds will also be discussed.

1.2 Background

1.2.1 Lichen

The term ‘lichen’ refers to a symbiotic association between a filamentous fungus, the ‘mycobiont’, and at least one photosynthetic organism, the ‘photobiont’, consisting of a microalga, a cyanobacterium, or both (De Priest 2004). Eighty-five percent of all lichen-forming fungi associate with eukaryotic algae (phycobiont) and another ten percent associate with prokaryotic cyanobacteria (cyanobiont). The remaining five percent contain both a phycobiont and a cyanobiont within the same lichen thallus thereby forming tripartite lichens. The name lichens was coined in 1867, when Schwendener proposed his dual hypothesis of lichens (Honegger, 2000). The Schwender hypothesis at this time was rejected by many biologists because the common thought was that all living organisms were autonomous (Honegger 2000). Before Schwendener (1829-1919) discovered the dual nature of lichens, most biologists mistakenly treated lichens as plants and thought they were one organism (Honneger 2000). Hawksworth (1984) described the lichen as a stable, self-supporting, mutualistic symbiosis involving a fungus (the mycobiont) and a microalga and/or cyanobacterium (the photobiont). This theory was finally accepted after microscopy revealed the association between algae and fungi. Lichens represent more than 40% of all known Ascomycota species which also represent the dominant life forms in about 8% of the land surface of the earth, including polar, alpine and coastal habitats where fog and water vapor are abundant. The name of a ‘lichen species’ refers to the name of the fungal species. Around 100 photobiont species have been reported to be as-

sociated with more than 13,500 lichen-forming species, so that many different mycobiont species share the same photobiont (Friedl and Büdel 1996). The fungal partner in the lichen symbiosis is generally considered to be ecologically obligate (Honegger 1998). This means that although the algal and fungal partners may be cultured independently, both must be growing together to produce the distinct lichen morphology that is observed in nature (Honegger 1996). Lichens have the ability to live in harsh weather and can colonize a large variety of substrates. Also they can be found in very cold and dry environments, either at polar latitudes or at extreme altitudes. They can grow on or inside rocks, on or inside the bark of woody plants, on wood, soil, mosses, leaves of vascular plants, dry stone walls and on other lichens. Despite this extreme range of ecological adaptations, most lichens are sensitive to changing their preferred ecological conditions and have difficulty growing in non-native habitats. Wherever lichens grow abundantly in nature, they often add a colourful aspect to their habitat (Convey, P *et al*., 2014, Manenti, R, 2014, Hauck M. *et al*., 2014, Shrestha G. and Clair L., 2013, Whittet, R *et al*., 2013). The distinct colours of many lichens are due to the accumulation of diverse secondary compounds (Shukla, P *et al*., 2014).

1.2.2 Lichen secondary Metabolites

Lichens are gaining interest in pharmaceutical research as a source of novel secondary metabolites. These secondary metabolites are produced in the fungal partner of the lichen association, which utilizes the carbon fixed by the algal partner (Molnar *et al*., 2010). These secondary metabolites are comparatively small, but also chemically complex mol-

ecules. Beside the externally visible crystallized and non-crystallized pigments that are deposited in the upper surface layers of the lichen, also colourless substances are common, which are predominantly found in internal parts of the thalli. According to Asahina and Shibata (1971) and Dayan & Romagni (2001), the lichen compounds may be classified into the following groups: (1) aliphatic lichen substances (including acids, zeorin compounds, polyhydric alcohols); (2) aromatic lichen substances (including pulvic acid derivatives, depsides, depsidones, quinones, xanthone derivatives, diphenyleneoxide derivatives, nitrogen containing compounds, triterpenes, tetronic acids); and (3) carbohydrates (polysaccharides). Most secondary lichen metabolites are derived from the acetyl-polymalonyl pathway, while others originate from the mevalonic acid and shikimic acid pathways (Figure 1.1).

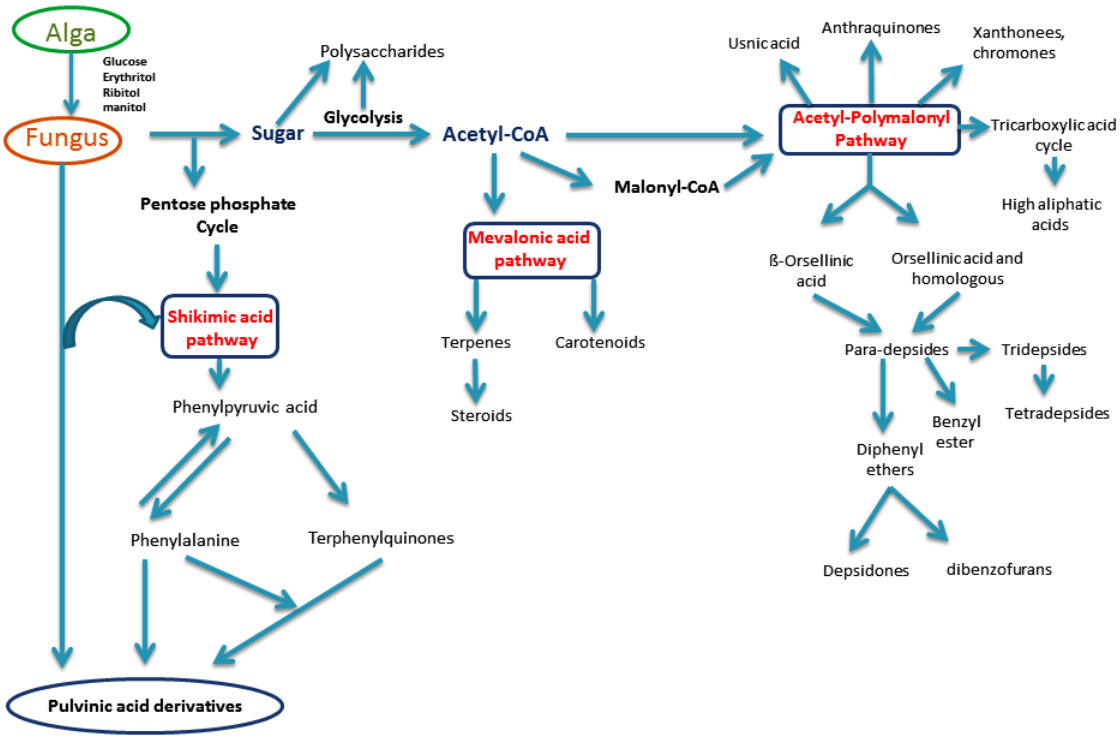


Figure 1.1 Different metabolic pathways that lead to the production of lichen secondary metabolites. Adapted from Elix (1996).

The distribution patterns of secondary metabolites are usually taxon specific, and therefore have been widely used in lichen taxonomy and systematics. One example is *Cladonia uncialis* lichen which has two chemotypes based on the secondary metabolites produced (Shrestha G. and Clair L., 2013, Zambare V.and Christopher L., 2012, Hawsworth 1976 and Culberson 1969). Lichen secondary metabolites are usually insoluble in water and can be extracted by using different organic solvents. The amount range from 0.1 to 10% of the dry weight of the thallus and sometimes they can reach up to 30% of the dry weight (Burkin A and Kononenko G., 2014, Zambare V.and Christopher L., 2012

And Varita, 1973). Lichen secondary metabolites have a wide array of roles that they can perform including protection from different consumers (Boustie J. and Grube M., 2005). The roles include volatile attractants for the same or other species, defense against pathogens, pests, and abiotic stresses (Zambare V.and Christopher L., 2012). In most cases the role of secondary metabolites in a particular organism, or the benefit, they provide is not yet known. For example, lichens that grow in extreme conditions have huge oxidative stress (e.g. Antarctica lichens) and produce large amounts of antioxidant substances when compared to other lichens that live in tropical places. Also, some lichens use a number of strategies to protect the light-sensitive algal partner against high levels of light and the damaging effects of UV radiation, for example, usnic acid produced by lichens may provide UV-B protection. Some of the lichen secondary metabolites are also associated with unpleasant taste (e.g. usnic acid), this represents another way for lichens to defend themselves from grazing by herbivores (Zambare V.and Christopher L. and 2012, Hauck M. *et al*., 2009).

1.2.3 Polyketides

Polyketides are remarkably diverse natural products that are present in a wide array of organisms such as bacteria, fungi, plants, insects, mollusks and sponges (Chen G. *et al*., 2014, Amnuaykanjanasin A. *et al*., 2005). The importance of polyketides comes from their wide range of pharmaceutical properties, including antibiotic, chemotherapeutic, and antioxidant activity (Wang Y *et al*., 2011). The wide spectrum of activities of polyketides makes them economically, clinically and industrially sought after molecules (New-

man, D. J., 2003). From cholesterol lowering drugs such as lovastatin, to anticancer drugs such as doxorubicin as seen in Figure 1.2, there are more than 40 FDA approved polyketide-based pharmaceutical drugs (Bochart, J. *Modern Drug Discovery* 1999, *2*, 22.). There is a great potential to discover novel polyketides by directed mutagenesis and combinatorial biosynthesis or having different analogues from a known polyketide following successful sequencing of PKS genes.

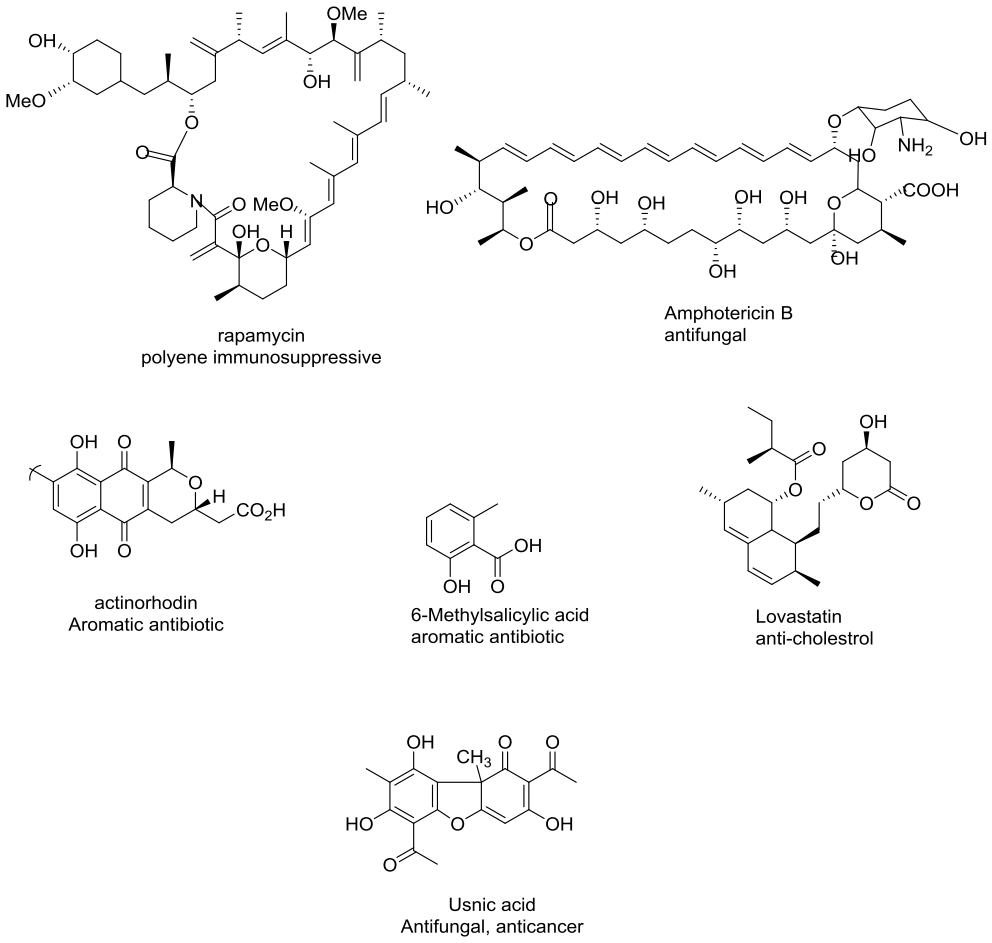


Figure 1.2 Examples of polyketide natural products.

1.2.4 Polyketides synthases

PKSs follow the same logic of assembly dictated by a set of catalytic domains similar to those found in fatty acid synthase (FAS) enzymes. Polyketides are synthesized by polyketide synthases (PKS) which consist at minimum of starter unit acyl-carrier protein transacylase (SAT), the keto-synthase (KS), the acyltransferase (AT), the product template (PT), the acyl carrier protein (ACP), and the thioesterase (TE) domains.

Polyketide biosynthesis is initiated by condensation of a simple acyl thioester starter with an extender unit. The starter unit for FASs is an acetyl moiety and malonyl-CoA is the extender building block, whereas PKSs can often utilize a variety of starter units including acetyl-, ethyl-, propionyl-, and butyryl-CoA or their equivalents, while the extender unit selected by the acyltransferase (AT) domain is malonyl-, methylmalonyl-, or occasionally ethylmalonyl-CoA. Chain elongation occurs through several cycles of condensation and modification reactions (Chooi Y. and Tang Y. 2012). The starting unit used by the SAT domain varies from PKS to PKS. Some systems, like PKS-NRPS SAT shuttle an NRPS product to the ACP domain to begin polyketide elongation. β- ketosynthase (KS) domain catalyzes a decarboxylative condensation to produce a ketoacyl thioester intermediate which can undergo successive reduction, dehydration and enoyl-reduction reactions, catalyzed by ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) domains respectively, to transform the β-keto group into a methylene group of the fully reduced fatty acid chain that is ultimately released by a thioesterase (TE) domain (Maier, T *et al.*, 2006). The presence of the reducing domains in the fatty acid synthase is mandatory while in the PKS genes are optional.

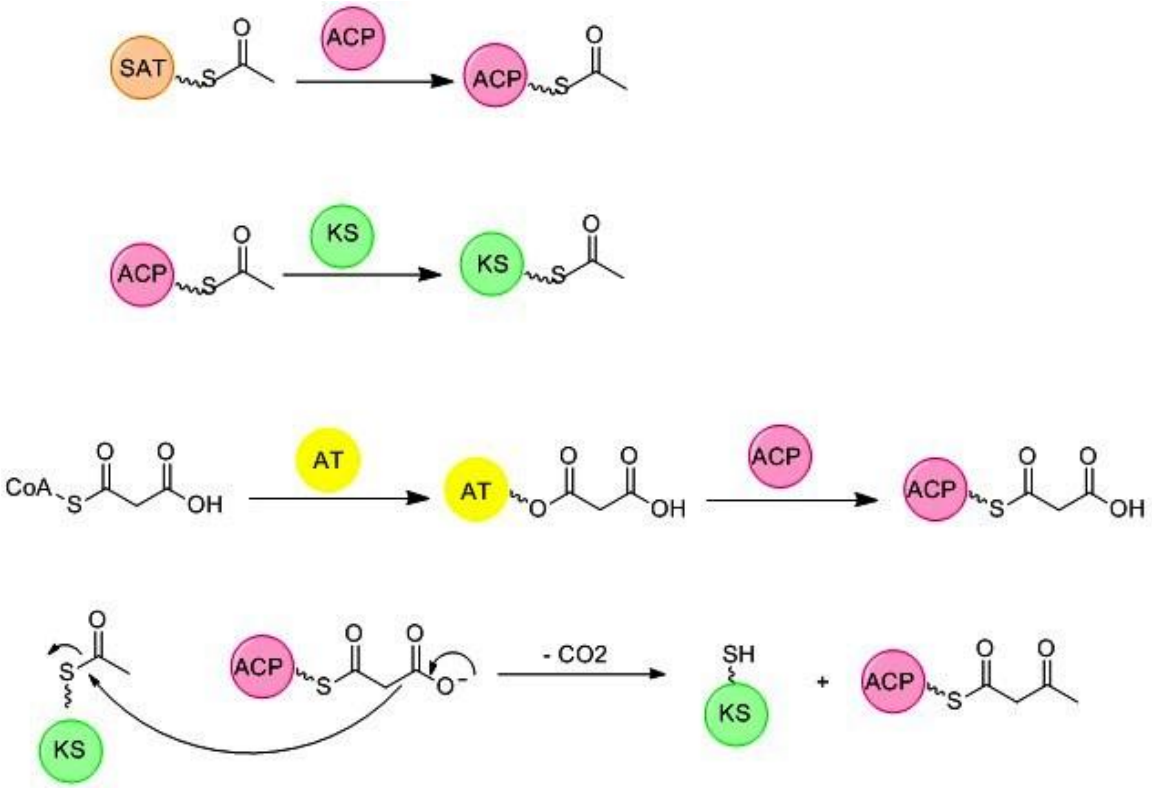


Figure 1.3 Loading and Elongation of a Polyketide Chain showing1-Acyl transacylase (AT) loads the starter unit on to the acylcarrier protein (ACP), 2-The ketosynthase accepts the starter unit from the ACP, 3-AT transfers the extender unit on to the ACP, 4-KS catalyzes condensation between the starter and extender units. PKSs can carry out partial or full reduction, raising enormous biosynthetic diversity. Adapted from-Hertweck, C. (2009).

PKSs are usually divided into three different types. Type I and type II, analogous to the FAS classification and type III PKS (Chalcone synthase) mainly present in plants. Like the type I FAS, type I PKS is a multifunctional enzyme which is present in both bacteria and fungi. The enzymes are arranged in modules and catalyze condensation reactions non-iteratively, according to the module function (Shen 2003). Each type I PKS module consists of several domains with defined functions, separated by short linker regions. Hence, the length of the polyketide carbon chain and the sequence steps by which it is built are determined by the number and order of modules in the polypeptides. Erythromy-

cin A and rifamycin are excellent examples of polyketides produced via modular type PKSs gene (August *et al*., 1998, Cortes *et al*., 1990). In an iterative Type I PKS which is mainly present in fungi, as for example the polyketide synthase responsible for lovastatin biosynthesis, the same module repeatedly catalyzes chain elongation and modification. In fungal polyketides, synthesis is initiated when the SAT domain transfers a starter unit onto the ACP domain, which then passes the starter unit to the KS domain. Next, the AT domain transfers a malonyl group from malonyl-CoA to the ACP. Finally, polyketide elongation is catalyzed by the KS domain, which catalyzes decarboxylative Claisen condensation to synthesize an elongated chain (Figure 1.3) (Keatinge-Clay A., 2012).

Type II PKSs; resemble the type II FASs in forming multi-enzyme subunits. Each PKS contains a 'minimal' set of core subunits. Two ß ketosynthase subunits KSα and KSß, an acyl carrier protein (ACP), and possibly a malonyl-CoA:ACP transacylase that is required for in vivo polyketide biosynthesis (Moore & Piel 2000). Tetracenomycin C is a polyketide produced by this family of PKSs (Gramajo *et al*., 1991). Type III PKSs are also called chalcone synthases (CHSs). These kinds of enzymes are simple homodimers of ketosynthases which catalyze the condensation of extender units onto a starter unit through iterative decarboxylative Claisen condensation reactions (Abe & Morita, 2010). Type III is characterized by the use of a single multifunctional KS-like enzyme active site with the capacity to incorporate acyl CoA extender units and perform all the necessary steps to assemble the polyketide backbone of the defined and proper length (Meslet-Cladière L. *et al*., 2013). Type III PKSs were previously thought to be only in plants or specific classes of bacteria, but have also been found in fungi and lichens (Lucia M. *et*

al., 2010, Nakano C. *et al*., 2009, Funa *et al*., 2007). Type iterative type I PKS genes are classified according to the presence or absence of the reducing domains with non-reducing PKSs being characterized by a lack of reducing domains and consisting minimally of a KS, ACP, AT and a termination domain. Partially- or highly-reducing PKSs may possess a combination of some or all of the reducing domains, as exemplified by fungal 6-methylsalicylic synthase (MSAS), which has a KR and DH domain (Child J, *et al*., 1996, Schmitt I *et al*., 2008). The classification of highly- vs. partially-reducing PKS genes is usually made based on the chemical structure of the product, with fungal metabolites such as lovastatin (Figure 1.2) (Campbell D *et al*., 2010) being typically classified as a highly reduced fungal metabolite.

1.2.5 Lichen polyketides

Lichen polyketides are produced by the fungal partner and no polyketide has been detected before from the photobiont partner. Most lichen polyketides are known to be unique to lichens and only a small number of these metabolites have been reported in other plants and fungi (Boustie J., 2011, Elix A., 1996). Although lichens have been commonly used for traditional medicines and cosmetics, commercial applications of lichens metabolites has been discouraged by the slow rate of growth of lichen, prohibiting mass-scale harvesting of their metabolites (Miao V. *et al*., 2001). With the need of new medical applications such as novel antibiotics, lichens and their unique metabolites have became the core interest to many of the research groups. Most of these unique lichen compounds come from polyketide pathways, and many are formed by intermolecular coupling of mostly

two monoaromatic units through ester, ether or carbon-carbon linkages. The coupled phenolics are divided into depsides, depsidones, depsones, biphenyls, diphenyl ethers and dibenzofuran based on the different kinds of bonds that join the two aromatic units as seen in Figure 1.4.

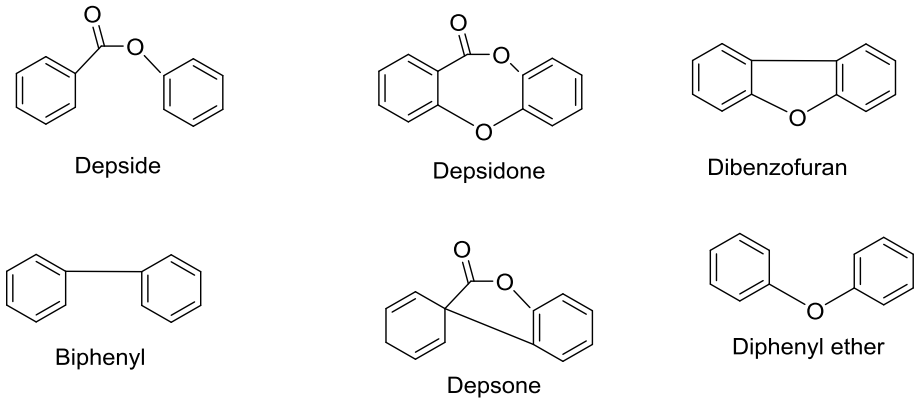


Figure 1.4 Different groups of coupled phenolics present in lichens polyketides where monoaromatic units are bonded by ester, ether or carbon-carbon linkages.

Depsones are the unique class that have been found only in lichens, while other different classes are occasionally found in the non-lichenized fungi (Huneck 2001; Huneck & Yoshimura, 1996). Monoaromatic units which represent the backbone for lichen polyketides are mainly orsellinic acid, β-orsellinic acid and methylphloroacetophenone (Figure 1.5). All of them are formed by a tetraketide chain via the polyketide pathway as seen in Figure 1.6.

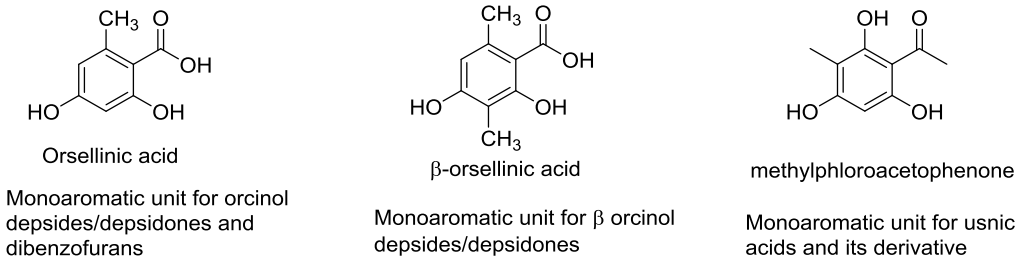


Figure 1.5 orsellinic acid, β-orsellinic acid and methylphloroacetophenone.

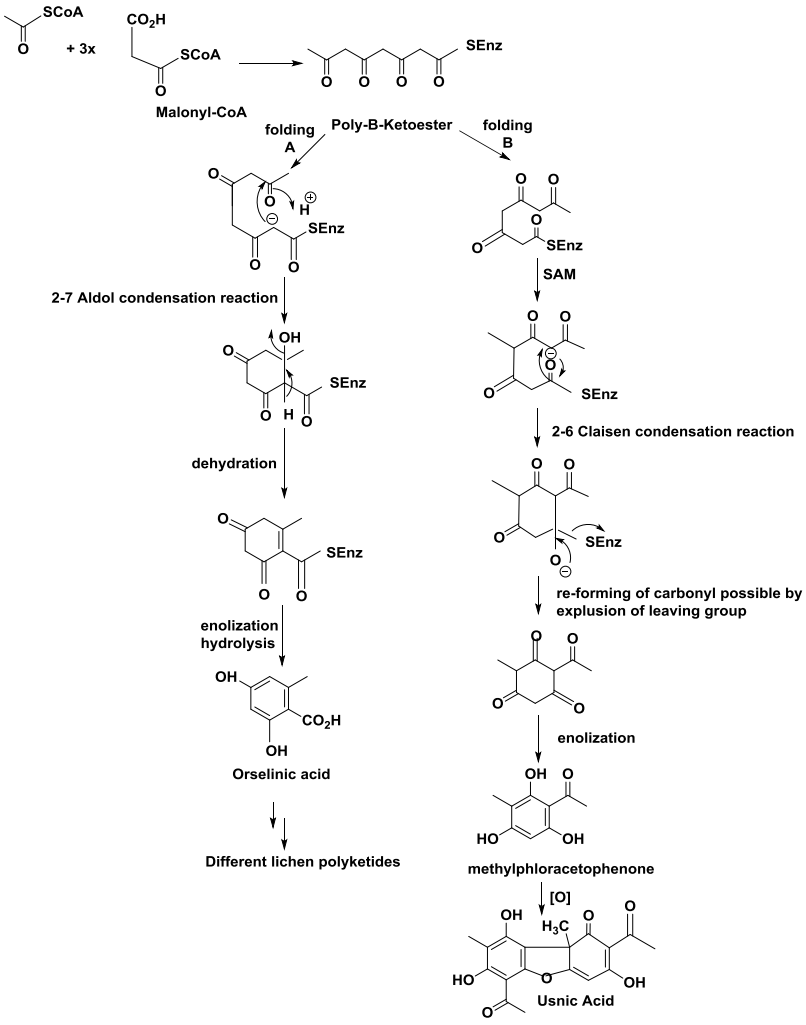


Figure 1.6 Proposed biosynthetic routes to orsellinic acid and methylphloroacetophenone via the polyketide pathway. PKS domains involved in the biosynthesis are KS, ketosynthase; AT, acyltransferase; ACP, acyl carrier protein, and CYC, Claisen cyclase. These monoaromatic units formed the basis of various coupled phenolics present in lichens. Figure is adapted from Dewick, P. (2006).

Polycyclic aromatic compounds in lichens such as anthraquinones and xanthones are also present in a wide range in different fungi and higher plants. One known example is the anthraquinone parietin which has been found in the vascular plant genera *Rheum*, *Rumex* and *Ventilago* and also in other fungi such as *Aspergillus* and *Penicillium* (Dewick 2002, Romagni & Dayan 2002, Thomas 2001, Peres *et al*., 2000). Also Averantin is produced by the lichen *Solorina crocea* and also by *Aspergillus* sp. (Figure 1.7). Resaerch in genetic engineering approaches such as functional heterologous expression of the biosynthetic genes is growing as prior successes in heterologous expression of biosynthetic enzymes in and from fungal species suggest this is a tenable means of harvesting polyketides produced by lichens (Wasil *et al*., 2013, Fujii, 2010, Cox, 2007).

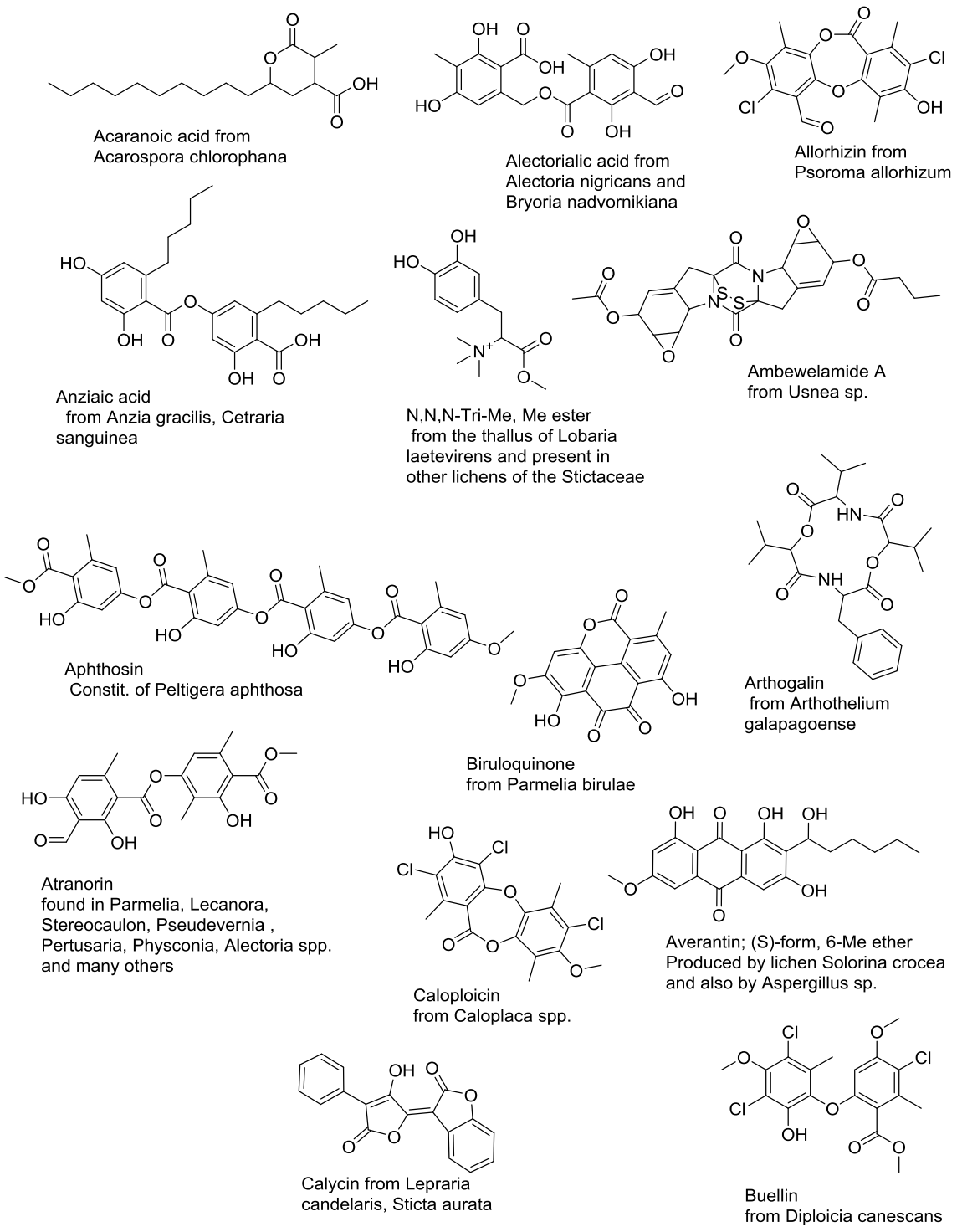


Figure 1.7 Examples of some lichen metabolites.

2 Chapter 2

GENERAL MATERIAL AND METHODS

2.1 Lichen material

The lichen sample of *Cladonia uncialis*, collection number Normore 8774, was collected in Manitoba, Canada, on 11 October 2008 in a Jack Pine dominated south-facing granite ridge (N54º 42' 24.7"; W101º 33' 53.1"). The lichen material was air dried and stored in the University of Manitoba Herbarium cryptogam division (WIN-C) until.

2.2 Liquid Chromatography

Cladonia uncialis lichen thallus was extracted with acetone. Acetone extract was then weighted and dissolved in acetone to bring the final concentration for the extract to 1mg/mL. HPLC analysis was carried out on Waters HPLC Separations Module 2695, combined with a PDA Detector Model 2996. The column was a μBondapak® Waters C18 (3.9 X 300 mm) with a column particle diameter of 15-20 μm, and with 125 Å pores. The flow rate was 1 mL/minute. The eluent was monitored continuously at 210-600 nm, and HPLC traces were displayed at 254 nm. The gradient was held at 20 % methanol to 80 % water in 0.075 % aqueous trifluoracetic acid for 10 minutes then the percent of

methanol was linearly increased up to 80 % and held at that composition for 20 minutes followed by a linear gradient back to 20 % methanol for 10 minutes and held there for 10 minutes. The total run time was 60 minutes. Usnic acid standard (ChromaDex, USA) and squamatic acid standard isolated from *Cladonia sulphurina* (Normore 9705) using preparative High Performance liquid chromatography were used as references for this experiment.

2.3 Extraction of Total RNA, cDNA synthesis and reverse transcription

The dry lichen sample was rehydrated with distilled water and incubated in natural light at room temperature for two hours before performing the RNA isolation. Total RNA was extracted from whole lichen thallus using the RNeasy Plant Mini kit (Qiagen, Mississauga, Ontario, Canada) following the manufacturer's directions. Fresh thallus (100 mg) consisting of young tissue at the branch tips was first ground by using mortar and pestle in liquid nitrogen. Immediately before the tissue thaw, 450 µL of buffer RLT (Plant Mini kit) was added to the tissue powder in a 2 mL microcentrifuge tube. The sample was vortexed vigorously and then incubated for 3 minutes at 56 °C. The lysate was transferred to a QIAshredder spin column placed in a 2 mL collection tube, and centrifuged for 2 min at full speed. The supernatant was transferred to a new microcentrifuge tube. Exactly 0.5 volume of ethanol (96–100%) was added to the cleared lysate, and then mixed immediately by pipetting. The sample including any precipitate was transferred to an RNeasy

spin column placed in a 2 mL collection tube and then centrifuged for 15 s at 8000 x g (10,000 rpm). Exactly 700 µL from Buffer RW1 (Plant Mini kit) was added to the RNeasy spin column and then centrifuged for 15 s at 8000 x g (10,000 rpm) to wash the spin column. Exaclty 500 µL from Buffer RPE was added to the RNeasy spin column and then centrifuged for 15 s at 8000 x g (10,000 rpm) to wash the spin column membrane. The same step was repeated with centrifugation for 2 minutes at 8000 x g (10,000 rpm). The RNeasy spin column was placed into a new 2 mL collection tube and centrifuged at full speed for 1 minute to dry the column completely. The RNeasy spin column placed in a new 1.5 mL collection tube and 50 µL of RNase-free water was added directly to the spin column membrane and then centrifuged for 1 min at 8000 x g (10,000 rpm) to elute the RNA. RNA was quantified on a NanoDrop 2000C (Thermo Scientific spectrophotometer, Wilmington, USA) and 1 µg of RNA was treated with DNase I (Invitrogen, Burlington, ON, Canada) following manufacturer's instructions.

The ThermoScriptTM RT-PCR system (Invitrogen, Burlington, ON, Canada) was used to synthesize the cDNA using approximately 100 ng of template RNA. First strand synthesis was carried out with oligo dT primers following the manufacturer's instructions.

2.3.1 RNA control experiment

Controls were tested for contaminating DNA in the total RNA and for contamination in the reaction components. For more confirmation two sets of primers, PKS1F (5'TACGAAGCCTAGAAATGGCT3') and PKS2R (5'ACGTTTGGCAGTTTCCTGTC3') (Fontaine *et al.* 2010) which amplify almost 500

bp of KS domain and LC1 (5'GAYCCIMGITTYTTYAAYATG3') and ISATc (5'GCATAITCICCIAGRCTRTG3') (Schmitt et al., 2008) which amplify an approximately 1700 bp fragment spanning most of the KS domain, the linker region and half the AT domain were used for the control experiment. The control experiment was performed on the genomic DNA as positive control and total RNA extracted from *Cladonia uncialis*. PCR was performed using 1X phusion HF buffer (New England Biolab), 200 µM of each dNTP (Promega) 0·5 µM of each primer, 2 Unit/ µL of Phusion DNA Polymerase (New England Biolab), 3% DMSO, about (50 -100) ng of RNA or 50 ng of total DNA as positive control, and water to 50 µL. The PCR conditions were denaturation at 98 ºC for 1 minute, then 30 cycles starting with denaturation at 98 ºC for 10 seconds, annealing at 54º-56 ºC for 30 seconds, and extension at 72 ºC for 1 minute. The reaction ended with a longer final extension at 72 ºC for 5 minutes. The annealing temperature varied according to the primers used in the PCR reaction.

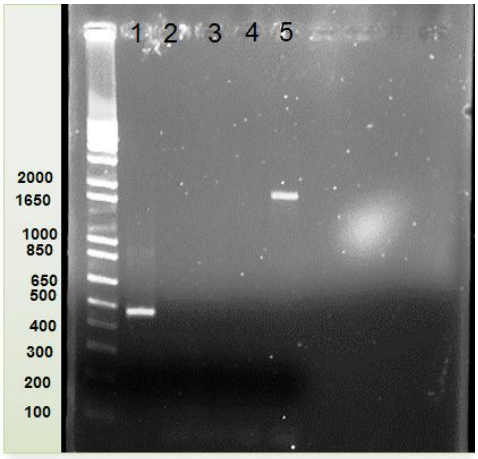


Figure 2.1 The results of the Control experiment. Lane 1:- Amplification product from the *Cladonia uncialis* genomic DNA by using PKS1F & PKS2R primers (first positive control). Lane 2:- Amplification product from *Cladonia uncialis* RNA (50 ng) using PKS1F & PKS2R primersLane 3:- Amplification product from *Cladonia uncialis* RNA using PKS1F & PKS2R primers (100ng). Lane 4 Amplification product from *Cladonia uncialis* RNA using Lc1F & JSATcR. Lane 5 Amplification product from *Cladonia uncialis* genomic DNA using Lc1F & JSATcR primers (second positive control).

2.4 DNA extraction

Total cellular DNA was extracted on the same podetia used for TLC using CTAB (Cetyl Trimethyl Ammonium Bromide). A modified protocol of Grube *et al*., 1995 was used for DNA extraction; the selected podetia were ground to a fine powder in 1.5 mL Eppendorf tubes using blue pestles. When sufficiently crushed, 500 µL of TES buffer [l00 mM Tris-HCI pH 8.0; 10 mM EDTA (ethylenediaminetetraacetate); 2% SDS (sodium dodecyl sulfate)] were added and the tubes were vortexed to homogenize the mixture. Sodium chloride (5 M; 1.4 M final concentration) and 10% CTAB (cetyltrimethylammonium bromide) (1% final concentration) were added. The tubes were vortexed briefly and placed in a 65 °C water bath for one hour. After this incubation period, the tubes were vortexed again and an equal volume of chloroform: isoamyl alcohol (24:1) was added. The tubes were mixed gently for 1 minute and centrifuged for 5 minutes at 5,000 rpm. The supernatant was collected and an equal volume of chloroform: isoamyl alcohol (24:1) was again added; the tubes were mixed gently and centrifuged for 5 minutes at 5,000 rpm. The supernatant was transferred to new Eppendorf tubes and 0.2 volumes of 5 M NaCl and 2.5 volume of 100% ethanol were added in order to precipitate the DNA. The tubes were mixed gently and left to stand at 4 ^{o}C for 20 minutes, after which they were centrifuged for 10 minutes at 13,000 rpm. The supernatant was poured off and the DNA pellet was washed with 80% cold ethanol and left to air dry. Once the pellet was dry, the DNA was resuspended in 50 mL sterile distilled water (sdH_2O) and stored in the freezer (-20 °C) until needed.

2.5 Gel electrophoresis

DNA extractions and PCR products were quantified using horizontal gel electrophoresis. The gels contained 1-1.5% agarose in 1X TBE (0.089M Tris, 0.089M boric acid, 2mM EDTA) buffer and were stained with ethidium bromide (0.5 mg/mL). Samples were placed in bromophenol blue (BP6) to facilitate loading and transferred to wells within the prepared gel. The gel was run at 120 V until the BPB was 1 cm from the bottom (approximately 30 min) and then examined under UV light with an Alphalnnotech 2200 Gel Document System (Fisher Scientific, Nepean, ON, Canada). DNA was quantified by comparing the intensity of the bands within the samples to the intensity of the 1650 bp band in the 1 kb Plus DNA Ladder (Invitrogen, Burlington, ON, Canada). The 1650 bp band contains 8% of total mass loaded. Therefore if 1 µg of 1 kb DNA ladder is added, then 0.08 µg of DNA is within the 1650 bp band. Bands of similar intensity will contain the equivalent amount of DNA. The 1 kb Plus DNA ladder, which ranges from 100 bp to 12,000 bp, was also used to determine the approximate length of the fragments generated by PCR.

2.6 DNA fragment purification

The DNA fragments were purified from the gel by using Wizard® SV Gel and PCR Clean-up system (Promega, Madison, USA). The gel slice that contained the DNA band of interest was dissolved in membrane binding solution (1µL for 1mg) and vortexed, then incubated at 50-65 ºC until the gel slice was completely dissolved. The membrane bind-

ing solution that contained the DNA band was then loaded onto a mini-column and incubated for 1 minute at room temperature, and the tube was centrifuged at 13000 rpm for 1 minute. The mini column was washed with 700 µL of membrane wash solution and then centrifuged as above. The DNA that bound within the mini-column was washed a second time with 500 µL of 'membrane wash solution' and centrifuged as above for 5 minutes. The purified DNA band was eluted from the mini-column in 50 µL of nuclease-free water and stored at -20 ºC.

2.7 Amplification of the ITS region

To confirm the identity of the natural lichen and cultured lichen fungi, the Internal transcribed spacer (ITS) region in the nuclear ribosomal repeat unit was determined by direct sequencing of the PCR product amplified with 1780F-5' forward primer (5'- CTGCGGAAGGATCATTAATGAG-3') (Piercey-Normore and DePriest 2001) which anneals to the 18S ribosomal small subunit and ITS4-3' reverse primer (5'TCCTCCGCTTATTGATATGC3') (White *et al*., 1990) annealing to the 26S ribosomal large subunit as shown in Figure 2.2. Amplification reactions for the natural lichen followed that of Beiggi & Piercey-Normore (2007) and were prepared in 0.2 mL PCR tubes on ice. Preliminary screening of each sample and amplification of ITS rDNA was conducted in 20 µL reaction volumes and contained approximately 10-50 ng DNA/reaction. Amplification for sequencing was carried out in eight 50 µL reaction volumes and contained approximately 25-100 ng DNA/reaction. A mastermix was prepared containing sterile distilled water (sdH_2O) to final volume, 1X PCR buffer (500 uM KCI,

100 µM Tris-HCl pH 8.3), 0.2 µM of each dNTP, 2.0 mM $MgCl_2$, 0.5 µM of each primer and 2 units of recombinant L DNA polymerase per sample. Recombinant L DNA polymerase was kindly provided by Dr. Loewen's laboratory (Department of Microbiology, University of Manitoba, MB, Canada). One drop of diluted DNA (containing 10-50 ng) was added to the bottom of each tube. The mastermix was briefly mixed and added to each tube. PCR reactions were carried out in a Techne Genius thermal cycler (Fisher Scientific, Nepean, ON, Canada). Touchdown PCR cycle consisted of initial denaturation conditions at 94 °C for 5 min, followed by two cycles of denaturation at 94 °C for 1 min, annealing at 60 °C for 1 min, and extension at 70 °C for 1.5 min. The annealing temperature was then decreased by two degrees after two cycles until reaching the final annealing temperature of 54 °C. Once the final annealing temperature was reached, 30 more cycles of denaturation, annealing and extension were performed. The size of the band was determined by using 1 kb Plus DNA ladder and horizontal gel electrophoresis. The PCR was repeated on a large scale by using 10 PCR tubes each containing 20 µL reactions. Amplified products to be used for sequencing were combined in a 1.5 mL Eppendorf® tube and precipitated using 0.2 volumes 5 M NaCl and 2.5 volumes 100% ethanol (ethanol). Samples were mixed gently and incubated at 4 °C for 20 min. They were then centrifuged at 13,000 rpm for 10 min and the supernatant was discarded. The remaining pellet was washed with cold 80% ethanol and allowed to air dry. Samples were then resuspended in 20 µL sdH20.

Figure 2.2 The binding sites and the extension directions for each primer used, for the 18S, 5.8S, 26S RNAs and two spacer regions.

2.8 Purification of the DNA fragments

The size, quantity, and purification of the amplified products was performed by horizontal gel electrophoresis in 1% agarose in 1X TBE buffer, stained with ethidium bromide, as previously described. The bands were visualized with a light box under short wave (254 nm) UV light, excised from the agarose gel, and placed in 1.5 mL Eppendorf® tubes and then purified using the Wizard SV Gel and PCR Clean-up system (Promega, Fisher Scientific, Nepean, ON, Canada), following the manufacturer's protocol as previously described (see Electrophoresis and Quantification of DNA extractions and purified PCR product). To quantify the DNA, the intensity of the band was compared with that of a known DNA quantity using a 1 kb plus DNA ladder (Invitrogen, Burlington, ON, Canada) as previously described in section 2.5. Sequencing reactions of the purified PCR product were performed in a total volume of 20 µL, consisting of 50 ng of purified DNA and BigDye 3.1 (Applied Biosystems, Foster City, CA, USA) following the manufacturer's instructions as described in section 2.9.

2.9 Cycle sequencing

Cycle sequencing reactions for all the purified DNA fragments and PCR products were set up in 0.2 mL PCR tubes on ice. Approximately 20-50 ng of purified DNA or purified PCR product, sdH_20 to a total volume of 20 µL, 3.2 pmol primer (one/reaction), 0.5X sequencing buffer and BigDye Terminator v3.1® (Applied Biosystems, Foster City, CA, United States), were contained in each 20 µL reaction volume. Cycle sequencing reactions were conducted in a Techne Genius thermal cycler (Fisher Scientific, Nepean, ON, Canada). Amplification conditions were initial denaturation at 96 °C for 2 min, followed by 25 cycles of denaturation at 96 °C for 10 sec, annealing at 50 °C for 5 sec, and extension at 60 °C for 4 min (according to manufacturer's instructions; Applied Biosystems, Foster City, CA, United States). Samples were then stored at 4 °C.

2.9.1 Post reaction clean up and sequencing

Following the cycle sequencing reactions, unincorporated fluorescent dyes were removed according to the manufacturer's instructions (Applied Biosystems, Foster City, CA, United States) by precipitating the DNA with 0.25 volumes 125 mM EDTA and 3 volumes 100% ethanol to each tube. Tubes were inverted four times to mix the samples, incubated at room temperature for 15 min, and then centrifuged at 4,000 rpm for 45 min. The supernatant was pipetted off and 80% cold ethanol was added to rinse the pellet. Tubes were then centrifuged at 4,200 rpm for 15 min and the supernatant was pipetted off again. The samples were then placed in the DNA120 SpeedVac (Thermo Savant, Holbrook, NY, USA) for 20 min to dry. To denature the final product for sequencing, 20

µL of Hi-Di Formamide (Applied Biosystems, Foster City, CA, United States) was added to each tube and incubated at 95 °C for 5 min. Following this incubation, tubes were immediately placed on ice before being loaded into a Genetic Analyser 3130 (Applied Biosystems, Foster City, CA, United States) for sequencing.

2.10 pETite N-His SUMO Kan vectors DNA

The pETite N-His SUMO Kan vector DNA (Lucigen) represents one of the improved alternatives to the most common pET expression vectors. The target gene was cloned under the control of the strong bacteriophage T7 promoter and the expression induced by a T7 RNA polymerase provided by the host cell. In general this plasmid carries the ribosome binding site; ATG, translation start site; stop translation end site, kanamycin resistance gene, repression of priming for low copy number; origin of replication, transcription terminators which prevent the transcription into or out of the insert, T7 terminator which follows the cloning site and then a 6XHis affinity tag which is fused to the amino terminus of the SUMO tagged protein. The vector is pre-linearized for directional cloning of the insert and also includes the 6XHIS-SUMO tag which has been shown to increase the yield and enhance the solubility for the majority of proteins.

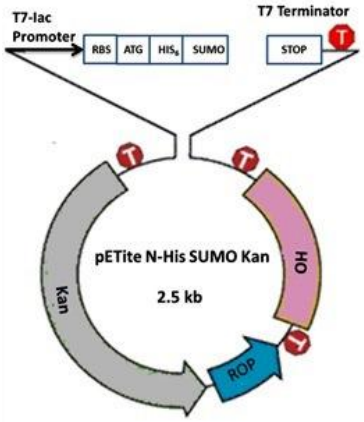


Figure 2.3 pETite N-HIS SUMO Kan expression vector. RBS, ribosome binding site; ATG, translation start site; HIS, represent histidine sequence Stop, translation end site; Kan, kanamycin resistance gene; ROP, Repressor of Priming; Ori, origin of replication. CloneSmart®transcription terminators (T) prevent transcription into or out of the insert, and a T7 terminator follows the cloning site.

2.11 HI control *E.coli* BL2 (DE3)

Escherichia coli BL21 strain is the common host for protein expression by using the pET vector. It has λDE3 lysogen phage which carries the gene for the T7 RNA polymerase. The T7 promote expression system is normally capable of producing more proteins than any other expression system. BL21 comes from *E. coli* strain B and it is not a recA mutant. The recA mutation keeps *E. coli* hosts from recombining homologous sequences of DNA in genomic clones. Another reason that BL21 is a good expression host is that BL21 is deficient in two key proteases, lon and ompT. The lon protease is an intracellular protease that *E. coli* makes to degrade abnormal proteins such as the protein that we want to express. It degrades the protein before the cells are even lysed. OmpT is a protease that *E. coli* makes to degrade extracellular proteins. It degrades the protein after our cells are lysed. IPTG is necessary with this strain to inhibit the *lacl* repressor and to induce the expression of the T7 RNA polymerase.

2.12 Microbiological media used in chapter 7

2.12.1 Luria-Bertani (LB) liquid medium

10 g tryptone, 5 g yeast extract, 10 g NaCl were dissolved in 1 L of ultrapure water and then autoclaved at 121 ^{o}C for 20 minutes. The appropriate antibiotic Ampicillin (100µg/mL) or Kanamycin (30 µg/mL) was added prior to inoculation.

2.12.2 LB solid media plates

Tryptone 10 g, yeast extract 5 g, NaCl 10 g and 15 g of agar are dissolved in 1 L of ultrapure water and autoclaved at 121 ^{o}C for 20 minutes. Appropriate antibiotic (Ampicillin (100 µg/mL) or Kanamycin (30 µg/mL) was added after the media cooled down (40^{o}C). About 20-25 mL of medium was poured in a petri dish in sterile laminar flow. The agar was then left to solidify inside the laminar flow and then plate was stored at 4^{o}C.

2.13 Other Cloning vector

2.13.1 pGEM®-T Easy Vector

pGEM®-TEasy vector is a linearized vector with a single 3'- terminal thymidine at both ends. It is a high copy number vector which contains T7 and SP6 RNA polymerase promoters flanking a multiple cloning region within the α-peptide coding region of the enzyme ß-galactosidase. In successful cloning the insert inactivates the α-peptide and this allows the identification of the recombinant by blue and white screening. β-galactosidase

is a protein which is encoded by the lacZ gene of the lac operon in bacteria. In the blue and white screening, the host *E. coli* strain carries the lacZ deletion mutant which contains the ω-peptide, while the plasmids used carry the lacZα sequence which encodes the first 59 residues of β-galactosidase, the α-peptide (Langley, E. et al., 1975). Both of them are not functional by themselves. However, when the two peptides are expressed together, as when a plasmid containing the lacZα sequence is transformed into lacZ cells, they form a functional β-galactosidase enzyme. The presence of an active β-galactosidase can be detected by X-gal, a colourless analog of lactose that may be cleaved by β-galactosidase to form 5-bromo-4-chloro-indoxyl, which then spontaneously dimerizes and oxidizes to form a bright blue insoluble pigment 5,5'-dibromo-4,4'-dichloro-indigo. In the case of successful cloning the DNA will be inserted within the lacZα gene, thereby disrupting the gene and thus production of α-peptide to give white colonies

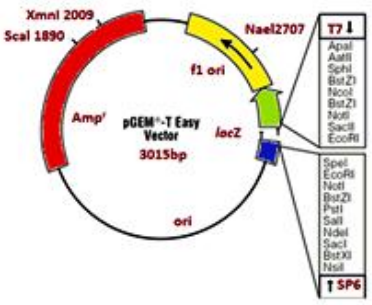


Figure 2.4 pGEM®-T Easy Vector Map.

3 Chapter 3

Identification of the Usnic Acid Biosynthetic Gene Cluster by de novo Genome Sequencing of *Cladonia uncialis*

3.1 Introduction

3.1.1 Usnic Acid

Lichen natural products are potential sources of pharmaceutically useful chemicals for humans. Clinical experiments have shown that some lichen natural products can be used as anticancer (Mayer *et al*., 2005), pain relief (Vijayakumar *et al*., 2000), and antipyretic drugs (Okuyama *et al*., 1995).

Polyketides are some of the most prevalent lichen secondary metabolites with a diverse array of biological activities resulting from the structural variation that occurs during bio-synthesis. One the most commonly occurring lichen polyketides is usnic acid which is

found in many lichen genera such as *Cladonia, Evernia, Usnea, Ramalina*, and *Alectoria*. Usnic acid was first described in 1844 and its biological activity has been subjected to extensive study (Ingolfsdottir, 2002). Usnic acid has been shown to have many biological activities such as antibacterial, antiviral, antioxidant, anti-inflammatory, and analgesic (Bessadottir *et al.,* 2012). In addition to all these activities usnic acid also possesses anticancer properties that was first reported for lung carcinoma and then demonstrated against a variety of human cancer cell lines (Shrestha and Clair, 2013). Despite all these features of usnic acid, its therapeutic application has not been introduced due to two major problems, its low water solubility and high hepatotoxicity (Francolini *et al*., 2013). It is also difficult to get usnic acid in large quantities using lichens because of the slow growth of the lichen and the difficulties in establishing pure fungal cultures (Fazio *et al*., 2012). To overcome these problems, alternative strategies must be developed for large-scale production of usnic acid and development of different analogues of usnic acid that retain its medicinal properties but have lower hepatotoxicity (Brunauer, 2009).

A critical breakthrough in our understanding of usnic acid production in lichens would be the identification of the biosynthetic gene cluster in the lichen fungal partner. This in turn would enable the alteration of the PKS gene via domain swapping or domain fusion thereby to produce usnic acid analogues that retain the beneficial bioactivity while mitigating the hepatotoxicity.

3.1.2 Proposed biosynthesis of usnic acid in lichen

Early work by Shibata in 1966 confirmed the polyketide origin of usnic acid as being derived from acetic acid (Taguchi H., 1966). A series of feeding experiments with radio-labelled (^{14}C) sodium acetate, diethyl malonate and sodium formate led to the isolation of radio-labeled usnic acid. Degradation experiments confirmed the ^{14}C atoms were incorporated at the expected positions in usnic acid as predicted by the polyketide biosynthetic pathway. This work also confirmed methylphloroacetophenone as the intermediate in the biosynthesis of usnic acid. Condensation between acetyl-CoA and malonyl-CoA to form acetoacetyl CoA, a diketide occurs via Claisen condensation. The tetraketide intermediate is bound to the polyketide synthase enzyme through a thioester linkage (Chapter 1, Figure 1.3). The tetraketide is then methylated by S-adeonsyl methionine (SAM). That methylation occurs prior to cyclization was demonstrated by Shibata when labeled phloracetophenone was not incorporated into usnic acid. This indicates that the methyl group on methylphloracetophenone must be incorporated prior to cyclization of the polyketide (Taguchi H *et al*., 1969). The cyclization of the methylated ß-keto ester is followed by enolization of the cyclohexatrione which results in a stable aromatic molecule intermediate, methylphloracetophenone (Figure 1.6).

3.1.3 Oxidation step in the proposed biosynthesis pathway of usnic acid

The final step in the biosynthesis of usnic acid is an oxidative homocoupling of two molecules of methylphloracetophenone catalyzed by a separate oxidative enzyme. After a radical carbon-carbon bond is formed between the two molecules, a Michael addition

forms hydrated usnic acid which undergoes dehydration to give usnic acid as seen in Figure 3.1. The binding site of the oxidative enzyme contains the two methylphloracetophenone molecules while the coupling takes place. Phenolic oxidative coupling can be performed by oxidase enzyme systems which are radical generators (Figure 3.1). These systems can perform oxidations or reductions using molecular oxygen as the electron acceptor. One known example of these systems is Cytochrome P450 dependent enzymes which require NADPH and oxygen as cofactors and can also perform one-electron oxidations of phenols (Hawranik D. et al., 2009, Stocker-Woergoetter, 2008).

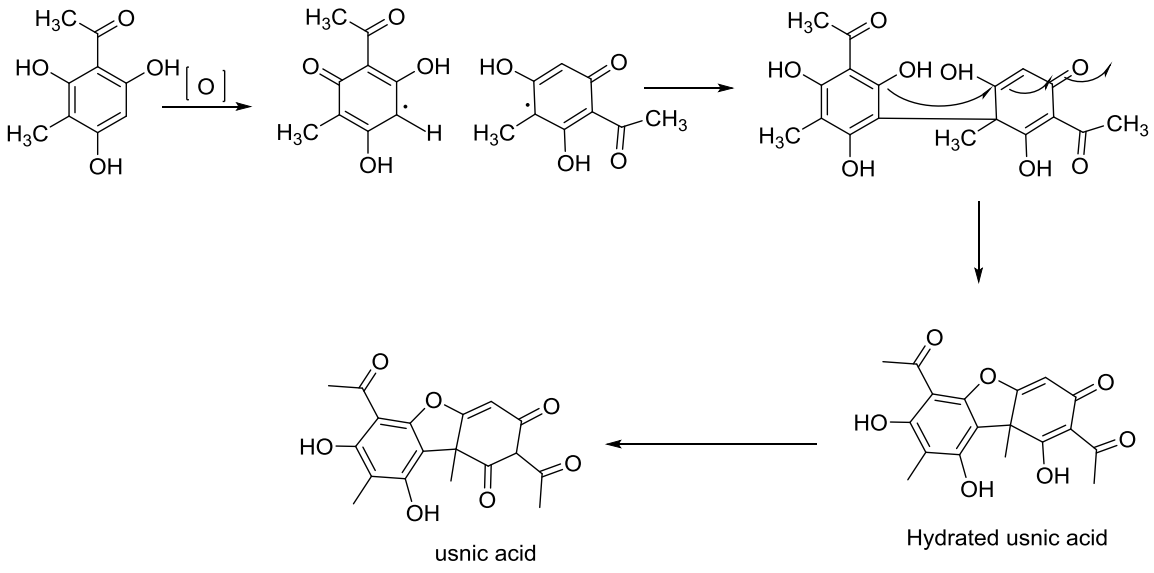


Figure 3.1 The homocoupling of two molecules of methylphloracetophenone by an oxidative enzyme to form usnic acid. Adapted from Dewick, P. (2006).

Based on the previous labeling experiments and the usnic acid biosynthesis pathway, the usnic acid gene cluster is predicted to be a fully nonreducing PKS gene methylphloracetophenone synthase (MPAS) with a single oxidative enzyme methylphloracetophenone oxidase (MPAO).

In this Chapter the identification of the usnic acid gene cluster through *de novo* sequencing of the genome of *Cladonia uncialis*, a known usnic acid producer is reported. This is the first example of the linking of a gene cluster to the production of a specific metabolite in lichenized fungi. This will help develop the technology required to harness the biosynthesis of usnic acid for the production of novel analogues with altered profiles of biological activity.

3.2 Materials and Methods.

3.2.1 Taxonomic identification of *Cladonia uncialis*.

The lichen, morphologically identified as *Cladonia uncialis* (Voucher number: Normore 8774, Herbarium, University of Manitoba, Winnipeg, Canada), was collected in October 2008, from Northern Manitoba on a south-facing granite ridge (N54º 42' 24.7"; W101º 33' 53.1"). The specimen was air dried and stored until needed (General material and methods chapter, section 2.1).

HPLC was carried out as described in section 2.2. The taxonomic classification of the lichen was proven by the amplification of the internal transcribed spacer (ITS) following the methods in section 2.6.

3.2.2 Culturing and genome sequencing of *Cladoina uncialis* mycobiont

The mycobiont was cultured from ascospores using the method of Yoshimura and colleagues (Yoshimura . *et al*., 2002) with modifications as described below: A single apotheum was washed and affixed to the lid of an inverted petri dish with petroleum jelly to permit spore release onto 1.5 % (w/v) water-agar media. Germinated spores were transferred onto solid malt-yeast extract (MY) media and maintained at 20 ^{o}C in the dark. The growing fungus was sub-cultured into fresh media every two months for one year. Total genomic DNA was extracted from the cultured mycobiont using the Gen JET Genomic DNA purification Kit (Thermoscientific, USA) following the manufacturer's protocol (Chapter 2, section 2.4). Confirmation of the identity of the mycobiont taxonomic classification was performed by amplification and sequencing by MICB (Manitoba Institute of Cell Biology, Cancer Care Manitoba, University of Manitoba) DNA sequencing services using 1780F (Piercey-Normore and Depriest, 2001) and ITS4 (White *et al*., 1990) primers (Table 3.1).

The lichen sample was *de novo* sequenced by MICB DNA sequencing services (Manitoba Institute of Child Health, University of Manitoba), using an Illumina MiSeq sequencer with a MiSeq Micro V2 sequencing kit at 43x coverage. The genomic DNA sample from *Cladonia uncialis* fungal culture was used to generate a paired-end DNA library for 150 bp paired-end sequencing reads.

3.2.3 De Novo Assembling and analyzing of *Cladonia uncialis* genome sequence

The raw data obtained from Illumina MiSeq sequencing were *de novo* assembled and analyzed into contigs using Seq Man Pro-software SeqMan NGEN program (NGEN) from Lasergene DNAStar (http://www.dnastar.com/t-products-seqman-ngen.aspx). Assembly was performed with 16 GB of RAM and was completed after 85 hours, 3 minutes and 27 seconds. This resulted in 2271 overlapping sequence read (contigs) of which 1859 are larger than 2 Kb.

The rapid *in silico* identification of putative polyketide synthase (PKS) genes was performed using Antibiotics and Secondary Metabolites Analysis Shell (antiSMASH 2.0) (http://www.secondarymetabolites.org/), an analytical program for identifying and annotating secondary metabolite biosynthetic gene clusters (Blin *et al*., 2013)

Uploaded FASTA sequences of assembled contigs were processed using the built-in HMMmer (Hidden Markov Models) tool that predicts and categorizes catalytic domains based on multiple sequence alignments of experimentally characterized signature proteins or protein domains. Probable domain architectures (e.g. PKS catalytic domains) were annotated based on the profile HMM library. The use of the antiSMASH smCOG tool (secondary metabolite clusters of orthologous groups) permitted prediction and categorical function of accessory post-PKS genes within the cluster (e.g. oxidative enzymes), based on all information available in the NCBI nucleotide database. The similarity of gene clusters to published sequences was evaluated through a built-in comparison tool based on BLAST sequence similarities.

3.2.4 Detection of expression of usnic acid PKS gene and its oxidative enzyme

The dry field-collected sample of *Cladoina uncialis* was rehydrated with distilled water and incubated in natural light at room temperature for two hours before RNA isolation. The total RNA extraction and the control experiment were carried out as described in Chapter 2. The synthesis of the 3' RACE-cDNA was performed on the total RNA extraction of the mycoboint and carried out according to the manufacturer's instructions (FirstChoice® RLM-RACE, Ambion, USA (Chapter 4, 4.2.2). The cDNA product was used as a PCR template, using the following reaction mixture: 1X phusion HF buffer (New England Biolab), 200 µM of each dNTP (Promega) 0.5 µM each of primers UAF2 and UAR2 (Table 3.1), 2 U/µL of Phusion DNA Polymerase (New England Biolab), 3% DMSO, 50 ng template cDNA, and enough water to total 50 µL. The PCR was run with an initial denaturation of 98 ºC for 1 minute, then 30 cycles starting with denaturation at 98 ºC for 10 seconds, annealing at 57 ºC for 30 seconds, extension at 72 ºC for 20 seconds, and a final extension of 72 ºC for 5 minutes. The expected size of the PCR product is 200 bp.

NAME	TARGET GENE	PRIMER SEQUENCE (5' TO 3')
UAF2	Usnic acid PKS	GCTTCTCCAGGTCGCTTATG
UAR2	Usnic acid PKS	AGCCAAAGTAATGCGACAGC
P450F1	Usnic acid oxidative	GTCTTCTGGGTTGCCATCAG
P450R1	Usnic acid oxidative	GAGACTTTGAAGGCGAGCAC
1780F	ITS	CTGCGGAAGGATCATTAATGAG
ITS4	ITS	TCCTCCGCTTATTGATATGC

Table 3.1 shows primers employed in experimental work on *Cladonia uncialis.*

3.2.5 Phylogenetic analysis of selected PKS domains

Sequences with similarities to the KS and CYC domains of the candidate usnic acid PKS gene that were publicly available in GenBank were used for sequence alignments using BLAST (Altschul *et al*., 1990). Nucleotide sequences converted to amino acid sequences were aligned with 100 available GenBank sequences for each domain using Clustalx V. 2.0. Maximum parsimony trees were produced using the option tree bisection and reconnection branch swapping. Heuristic searches were conducted using 1000 random addition replicates with a limit of 10 trees per search and bootstrap searches of 1000 re-samplings The sequence of the *Trichoderma atroviride* (EHK44445) PKS gene was selected as an outgroup for the KS domain tree while *Trichoderma virens* (EHK19347) hypothetical protein was chosen for the CYC domain tree.

3.3 Results and Discussion

3.3.1 Taxonomic Identification of *Cladonia uncialis*.

A fresh field-collected sample of lichen morphologically identified as *Cladonia uncialis* was examined for the production of usnic acid. An acetone extract of *C. uncialis* tissue was analyzed by HPLC (UV-DAD 200-600 nm) and only a single peak representing a single secondary metabolite was detected. This metabolite has a retention time (23.7 min) and UV spectra (λ_{max} 233 and 281 nm) identical to that of a standard sample of usnic acid (ChromaDex). There were no other signals detected in the HPLC trace of the acetone extract as in Figure 3.2.

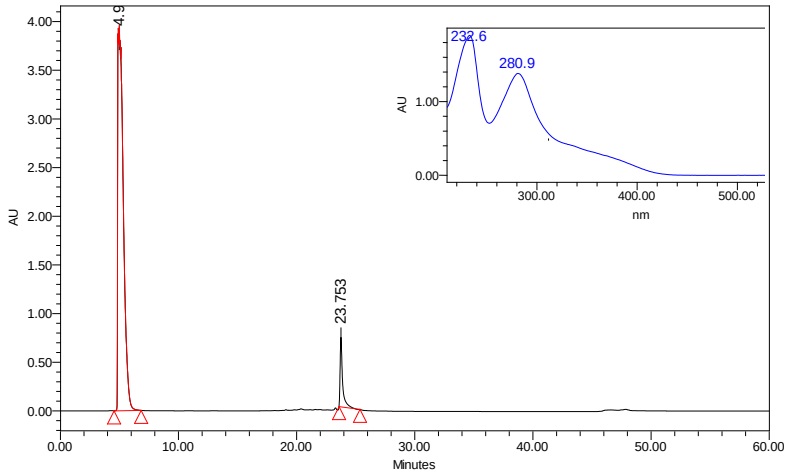


Figure 3.2 HPLC analysis of secondary metabolites produced by *Cladonia uncialis* with UV absorbance at 254 nm. Only usnic acid was detected.

The identity of the collected organism was confirmed as *C. uncialis* by amplification of the internal transcribed spacer (ITS) region of the nuclear ribosomal repeat unit. A BLAST search of the amplified sequence gave a match of over 95% identity (with 0% gaps) with ITS sequences from other deposited samples identified as *C. uncialis.*

3.4 *Cladonia uncialis* mycobiont genome sequencing

In order to carry out the whole-genome sequencing of *C. uncialis* it was necessary to first generate cultures of the mycobiont partner that were isolated from the algal partner. Culture conditions were developed according to Yoshimura *et al*., (2002), as described in section 3.2.2. To confirm the identity of the mycobiont, the ITS region from a sample from one of the mycelia mats was amplified and sequenced following a procedure identical to that described above. The resulting sequence was 100% identical to that generated for the ITS region from the field-collected sample of *C. uncialis,* therefore confirming the identity of the cultured mycobiont as originating from *C. uncialis.* For whole-genome sequencing the genomic DNA was extracted and isolated from 18-month old mycelial mats. Sequencing was carried out at a facility operated by the Manitoba Institute of Cell Biology using an Illumina MiSeq with SOLiD sequencing at 43x coverage. The raw DNA sequences were assembled *de novo* into 2271 contigs of which 1859 were greater than two kilobases in length. The assembled contigs were then further analyzed for the presence of secondary metabolite gene clusters as described in Section 3.2.3.

3.4.1 Identification of usnic acid gene cluster

The contigs were scanned for the presence of sequences that had similarity to polyketide synthase (PKS) genes using the Antibiotics and Secondary Metabolites Analysis Shell (antiSMASH 2.0). AntiSMASH is software that was designed to detect and characterize secondary metabolite gene clusters based on the sequence similarity to experimentally characterized sequences (Blin *et al*., 2013). The architecture of a PKS involves, at a minimum, four domains: a ketosynthase (KS), an acyl transferase (AT), an acyl carrier protein (ACP) and a terminal domain, usually a thioesterase (TE) which off-loads the PKS product via hydrolysis (Keatinge-Clay, 2012, Chan *et al*., 2009, Tsai, S.C *et al*., 2009). As described before in Chapter 1, type I PKSs are divided into non-reducing PKSs which lack reducing domains; partly reducing which have some reducing domains and highly reducing which contain all the reducing domains. The chemical structure of methylphloracetophenone suggests that the MPAS does not possess any reducing domains as indicated by the presence of three phenolic oxygens in the structure. In addition, in agreement with the earlier work of Shibata (1966), the presence of a MeT domain is also indicated, which would be responsible for the incorporation of the methyl group from SAM. The biosynthesis of methylphloracetophenone and the suggested domain architecture of MPAS is shown in Figure 3.3.

AntiSMASH was employed to scan the genome for the presence of individual genes with homology to predicted PKS genes. Our initial scans identified 32 candidate sequences in the *C. uncialis* genome that were consistent with PKS genes. However, out of these 32

PKS genes only 5 genes (Fig 3.4.) are nonreducing with a methyl transferase (CMet) domain that is a requirement for the biosynthesis of usnic acid (Figure 3.3)

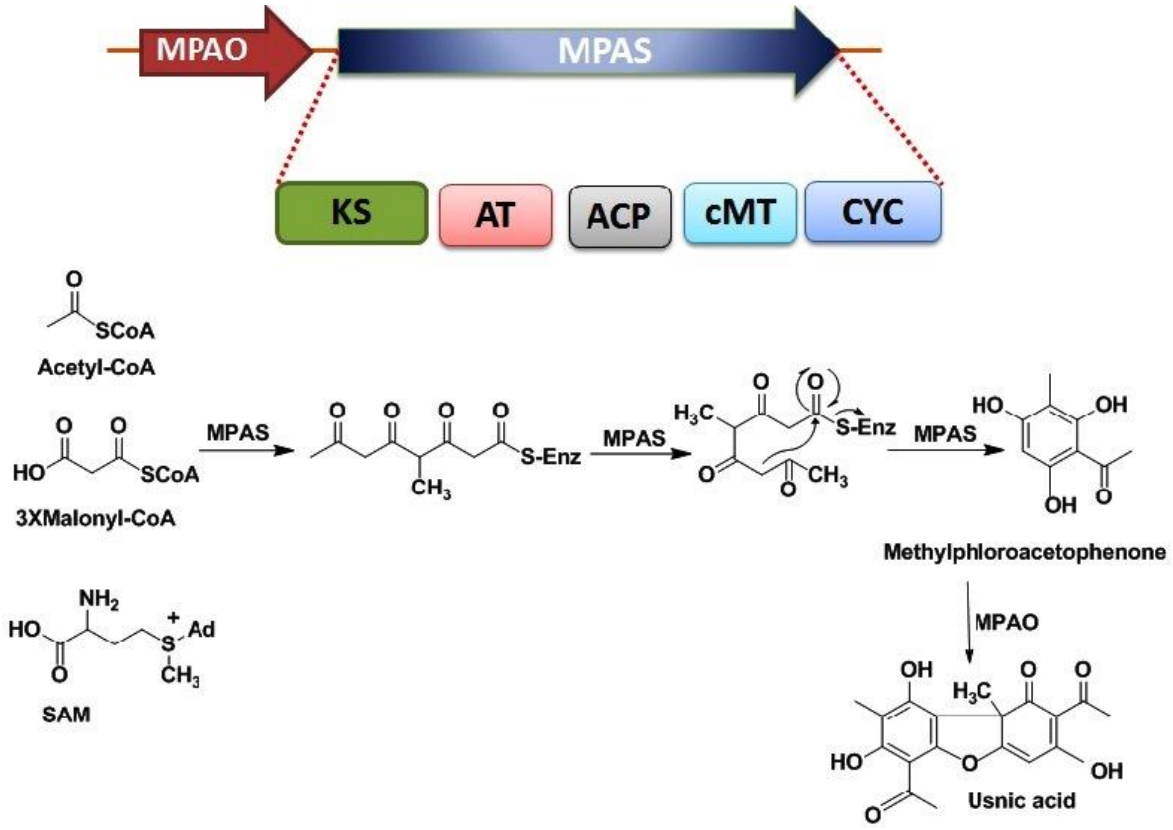


Figure 3.3 The usnic acid gene cluster and biosynthesis in *Cladonia uncialis*. Upper part shows the organization of the usnic acid biosynthetic gene cluster. Methylphloroacetophenone synthase (MPAS), encoded by *mpas*, is a non-reducing polyketide synthase possessing the following catalytic domains: starter acyl transferase (SAT), ketosynthase (KS), acyl transferase (AT), acyl carrier protein (ACP), a carbon-carbon methyl transferase (CMet), and a terminating Cyclase domain (Cyc). Methylphloroacetophenone oxidase (MPAO), encoded by *mpao*, is an oxidative enzyme characteristic of cytochroe p450s, found upstream of *mpas*. The lower part shows the biosynthesis of usnic acid by MPAS and MPAO. Three molecules of malonyl Coenzyme A and *S*-adenosylmethionine (SAM) are assembled by MPAS (primed with acetyl Coenzyme A) to form methylphloroacetophenone. Oxidative dimerization of two molecules of methylphloroacetophenone to complete the synthesis of usnic acid.

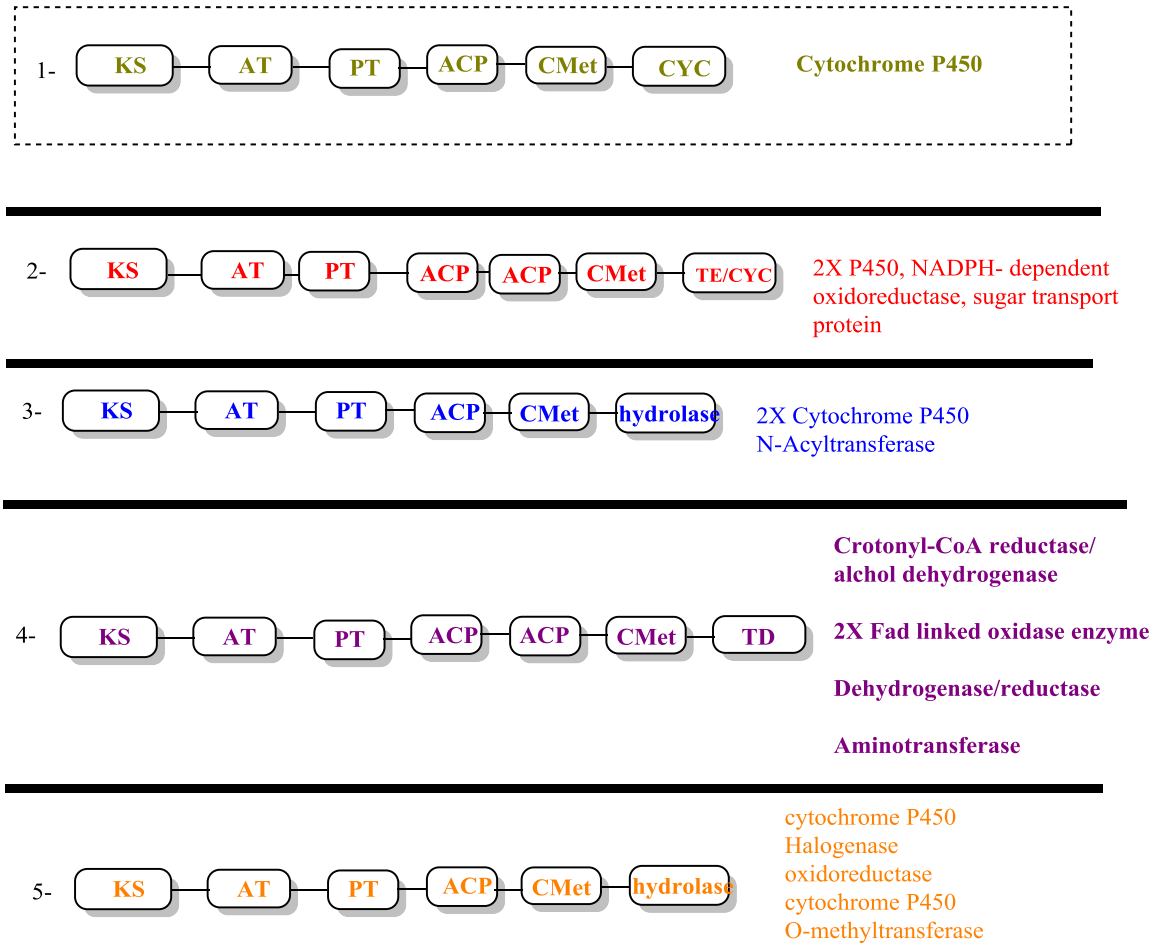


Figure 3.4 *Cladonia uncialis* type 1 non reducing polyketide synthase genes possessing a C-methyltransferase domain. Gene 2 has two cytochrome p450 genes that are closely linked, as well as an NADPH-dependent oxidoreductase. Gene 3 has 2X Cytochrome P450, N-Acyltransferase. Gene 4 has two FAD-linked oxidative proteins, crotonyl-CoA reductase/ alchol dehydrogenase, dehydrogenase/reductase and aminotransferase enzymes. Gene 1 represents the usnic acid PKS gene candidate possesing a single associated cytochrome p450 gene, consistent with oxidative dimerization of (methylphloracetophenone) into usnic acid.

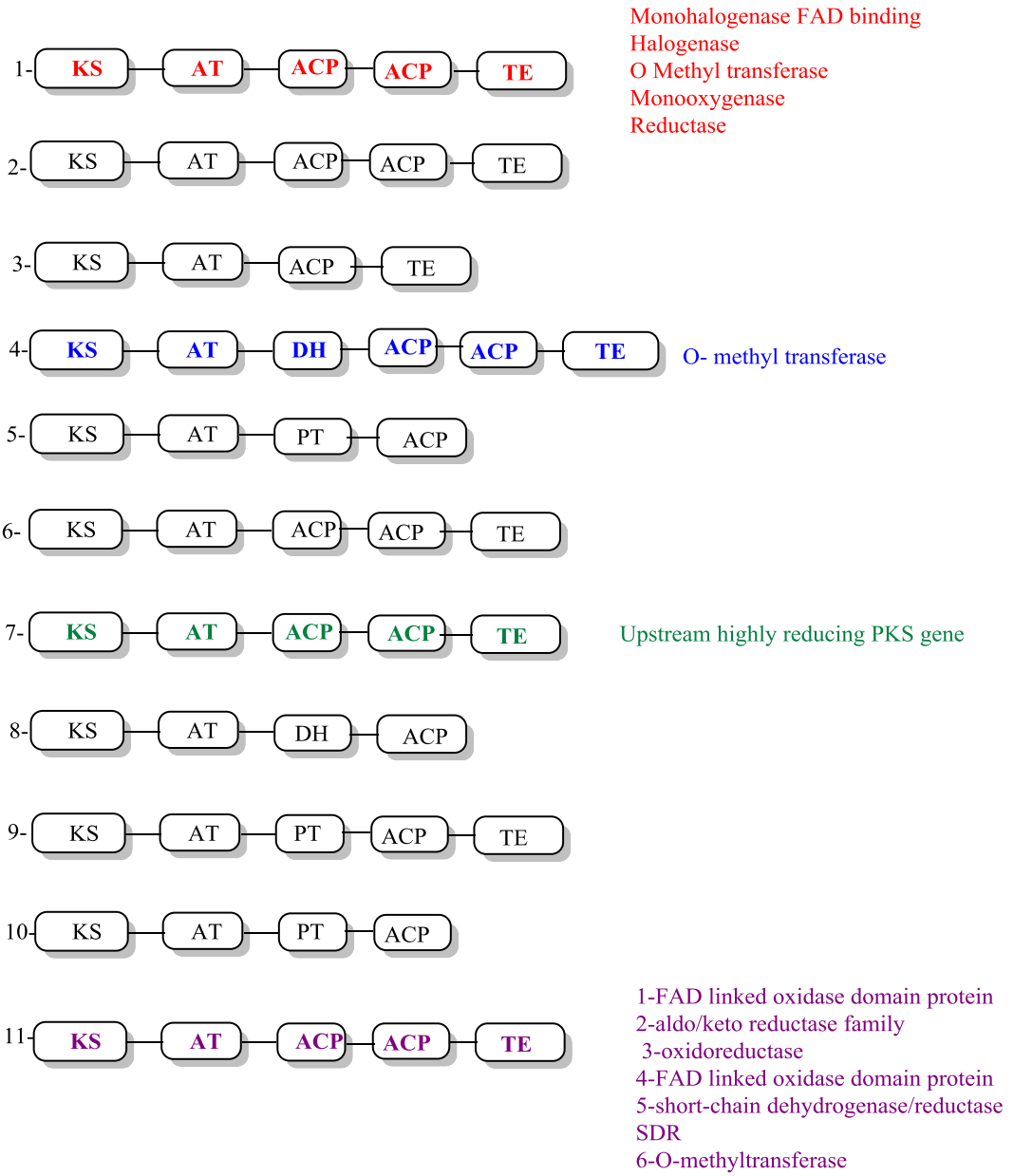


Figure 3.5 *Cladonia uncialis* type 1 non-reducing polyketide synthase genes with no C-methyltransferase domain.

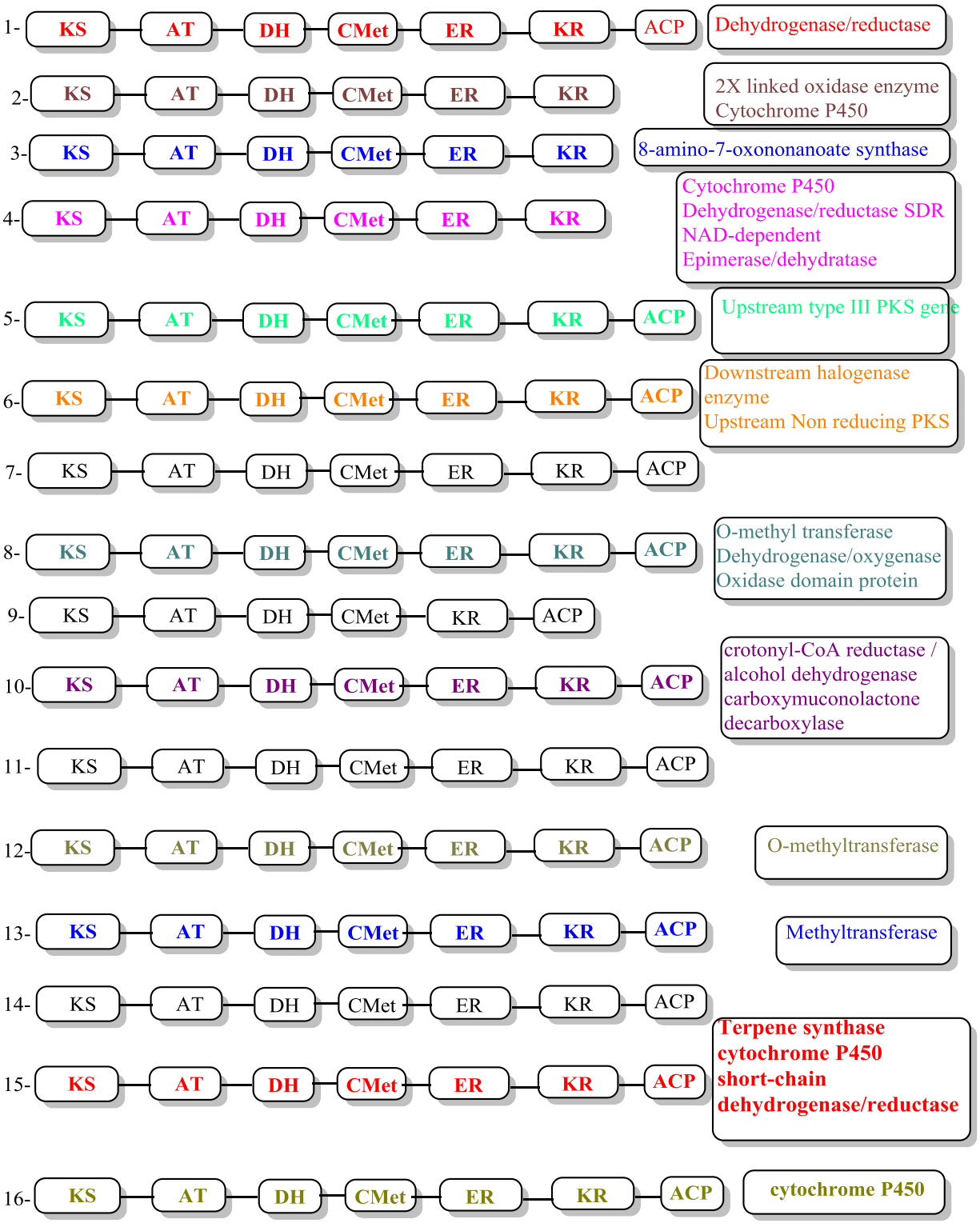
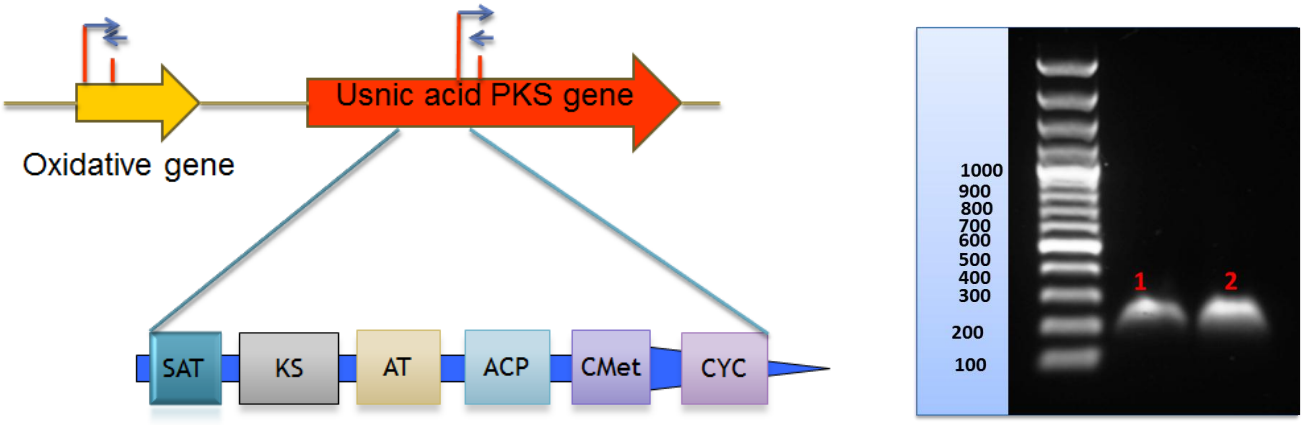


Figure 3.6 *Cladonia uncialis* type 1 reducing polyketide synthase genes.

Deductive analysis based on biosynthetic rationales permitted progressive elimination of all gene clusters in Figures 3.5 and 3.6. Based on the predicted architecture of the usnic acid gene cluster I was then able to narrow the search to five of the identified PKS genes (Fig 3.4). These five candidate PKS genes display domain architectures consistent with the lack of reducing domains predicted for *mpas*. In order to deduce which of these five remaining genes was *mpas*, the accessory genes associated with each PKS were examined. One of the genes (Gene 2 in Fig. 3.4) had two cytochrome p450 genes that were closely linked, as well as an NADPH-dependent oxidoreductase. Another gene (Gene 3 in Fig 3.4) had two Cytochrome P450, N-Acyltransferases. Gene 4 in Figure 3.4 had two FAD-linked oxidative proteins, crotonyl-CoA reductase/ alchol dehydrogenase, dehydrogenase/reductase and aminotransferase enzymes. These 4 genes are not consistent with usnic acid biosynthesis and only one candidate gene remained, possessing a single associated cytochrome p450 gene, consistent with oxidative dimerization of (methylphloracetophenone) into (usnic acid) (Gene 1 in Fig 3.4). These genes and protein products have hence been assigned as the PKS methylphloracetophenone synthase (MPAS), encoded by *mpas*, and methylphloracetophenone oxidase (PMAO), encoded by *mpao*. The presence of a single oxidative gene is consistent with the proposed biosynthesis of usnic acid where it seems most likely that a single enzyme is involved in the dimerization of methylphloroacetophenone. This is also consistent with earlier results that demonstrate that usnic acid can be formed from methylphloroacetophenone in a single enzyme-catalyzed oxidative step (Ingolfsdottir, 2002, Hawranik, 2009).

3.4.2 Transcription of the usnic acid gene cluster in *C. uncialis*

In order to confirm the transcription of *mpas* and *mpao*, the RNA extract from the field-collected sample of *C. uncialis* was examined to see if there is any amplification of the corresponding cDNA sequence. Using the sequence predicted from the genomic DNA, PCR primers specific for both *mpas* and *mpao* were generated in this study as shown in Table 3.1. The primers were specific for the KS region in *mpas* and for the central sequence of *mpao*. The primers were designed to amplify 200 bp from each gene, (UAF2, UAR2) to target the *mpas* gene and (P450F1, P450 R1) to target the *mpao* gene (Table 3.1). In separate experiments, bands of the predicted lengths were observed for *mpas* and *mpao* from the cDNA (Figure 3.7). Sequencing of these cDNA fragments confirmed they originated from both *mpas* and *mpao* especially in light of the fact that the only secondary metabolite detected in *C. uncialis* was usnic acid (Figure 3.2). This result provided additional evidence that the usnic acid gene cluster was indeed identified.



A **B**

Figure 3.7 A- The position of the gene specific primers that amplify 200 bp fragments of *mpas* and *mpao* genes. B- Electrophoresis of the cDNA of the usnic acid gene cluster.The first band represent 200bp from *mpas* P and the second band represents 200 bp from the *mpao*.

3.5 Phylogenetic relationship of usnic acid KS domain to other sequenced fungal KS domains

In order to get a picture of the relationship between *mpas* and other PKS modern genes, phylogenetic trees were constructed for two of the domains of this gene. The sequence from *mpas* that corresponds to the KS domain was chosen to be examined. There were two reasons for this choice. First, the KS domain is ubiquitous to all polyketide synthase genes, responsible for catalyzing the key carbon-carbon bond-forming Claisen condensation between malonyl-CoA and the growing polyketide chain. This ubiquity provides for the direct comparison of a broad selection of PKS genes, and in particular, fungal non-reducing types. The second reason is that in recent years a large number of sequences for the KS domains from fungal non-reducing PKS genes, particularly from lichen fungi, have been deposited in GenBank. This large number of deposited sequences allows us to more clearly trace the relationship between *mpas* and other PKS modern genes. However

the function of only a few of the deposited PKS gene fragments has been assigned. The *mpas* KS sequence was predicted to fall into a unique clade and potentially offer a tool to identify other *mpas* genes that have been sequenced from other strains of usnic acid-producing lichens. The majority of the phylogenetic analyses for the PKS genes were done on the KS domain assuming that the evolutionary relationship of the KS domain (as the most conserved domain present in all PKSs) would reflect the domain architecture of the particular PKS genes. Earlier PKS genes analysis work suggested that the segregation of the KS domain sequence is consistent with the non-reducing, partly-reducing and highly-reducing PKS grouping (Nicholson *et al*., 2001, Bingle *et al*., 1999). In 2003 Kroken *et al*., did an extensive phylogenetic analysis for a large number of PKS genes across several fungal genomes. At the time of this project the functions of most of these PKSs were uncharacterized. Kroken *et al*., (2003) did their analyses by using the predicted KS amino acid sequence. The analysis showed a significant correlation between the PKS phylogeny and domain architecture. The phylogenetic analysis of Kroken *et al*., (2003) and other related analyses (Varga *et al*., 2003, Nicholson *et al*., 2001 and Bingle *et al*., 1999) suggested that the KS domain phylogeny could potentially be used to determine the domain architecture of a new PKS and predict the type of the polyketide product if it is reducing or nonreducing. The phylogenetic tree constructed from the evolutionary analysis of *Cladonia uncialis* ketosynthase domain is shown in Figure 3.8. Almost all the PKS genes involved in the phylogenetic analysis have a CMet domain. This is consistent with the previous analysis and classification work performed by Kroken et al., (2003) which mentioned that the genes with the same domain architecture cluster together. The

DH domain was present in some of these genes (Figure 3.8). The *Cladonia uncialis* PKS gene matches by 100% with *Usnea longissima* PKS4 which is a known lichen species that produces usnic acid. The phylogenetic tree constructed from the evolutionary analysis of the ketosynthase domain is shown in Figure 3.8.

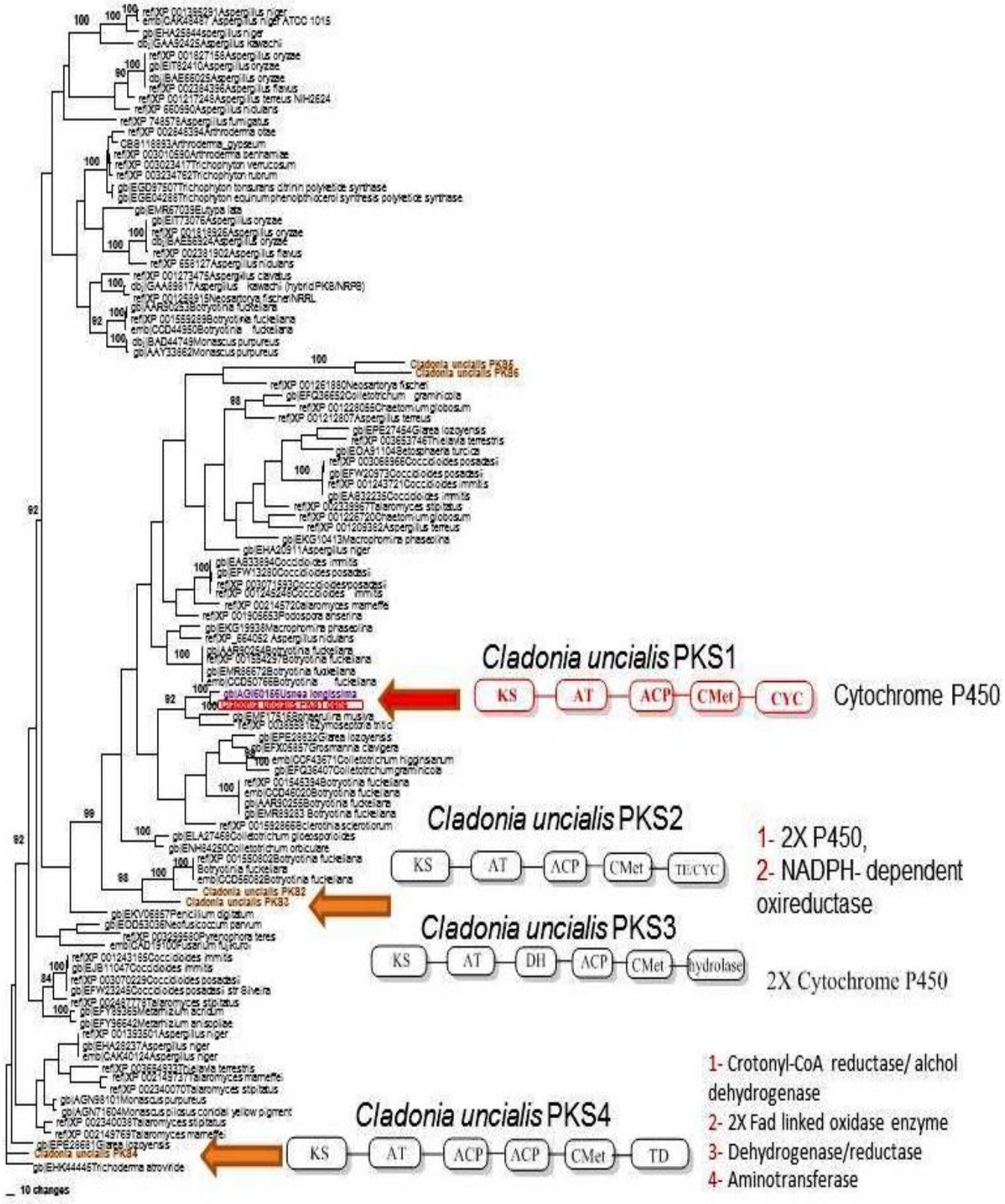


Figure 3.8 Phylogeny of the KS domains from predicted *Cladonia uncialis* usnic acid PKS gene and homologous fungal non-reducing type I PKSs. Please refer to Appendix A for a high resolution one.

3.6 Phylogenetic relationship of the CYC domain in usnic acid PKS gene to sequenced PKS genes.

The CYC domain of *mpas* is predicted to be unique to the biosynthesis of methylphloroacetophenone and therefore, unique to usnic acid. Usnic acid is the most common lichen secondary metabolite derived from methylphloroacetophenone, making the biosynthesis of methylphloroacetophenone a critical requirement for the biosynthesis of usnic acid**.** The key step in the biosynthesis of methylphloroacetophenone is the folding of a tetraketide such that a Claisen-type condensation leads to ring formation (Figure 3.3). The vast majority of other lichen metabolites appear to be derived from 3-methylorsellinic acid, which in turn is formed when the tetraketide chain folds up into an alternative aldol-type condensation, catalyzed by an aldolase (ALD) domain, leading to ring formation (Figure 1.6). It has yet to be demonstrated if this occurs with concomitant thioester hydrolysis. Owing to the potentially unique nature of the CYC domain, a phylogenetic tree was constructed to examine the evolutionary relationships that may be evident from this domain (Figure 3.9). Although there were fewer terminal PKS domains deposited in GenBank than for the KS domain, there were enough to make a meaningful comparison. It was predicted that the CYC domain may potentially fall into a unique clade, possibly grouping with other PKS genes that were responsible for the biosynthesis of methylphloroacetophenone. One hundred fungal and lichen PKS amino acid sequences obtained from GenBank were aligned with four CMet-possessing genes sequenced from *C. uncialis* (Gene 1-4 in Fig. 3.9), including the putative usnic acid PKS gene (Gene 1, Figure 3.9). Genes 5 and 6 (Fig. 3.8) were excluded because they were highly-reducing PKSs, and

were expected to group outside the clade. The PKS gene from *Usnea longissima* (Ascension No. AGI60156; Wang & Hur, unpublished GenBank submission), the same gene acquired from GenBank that grouped with the putative usnic acid PKS gene in the KS phylogenetic analysis, was the only gene that displayed grouping with the usnic acid gene sequenced from *C. uncialis.* These combined phylogenetic analyses provide supporting evidence for the identification of *Cladonia uncialis* PKS1 gene encoding the methylphloracetophenone-producing PKS leading to usnic acid, and suggests that the sequence of *Cladonia uncialis* PKS1 can be used as a marker for the identification of other *mpas* synthase genes.

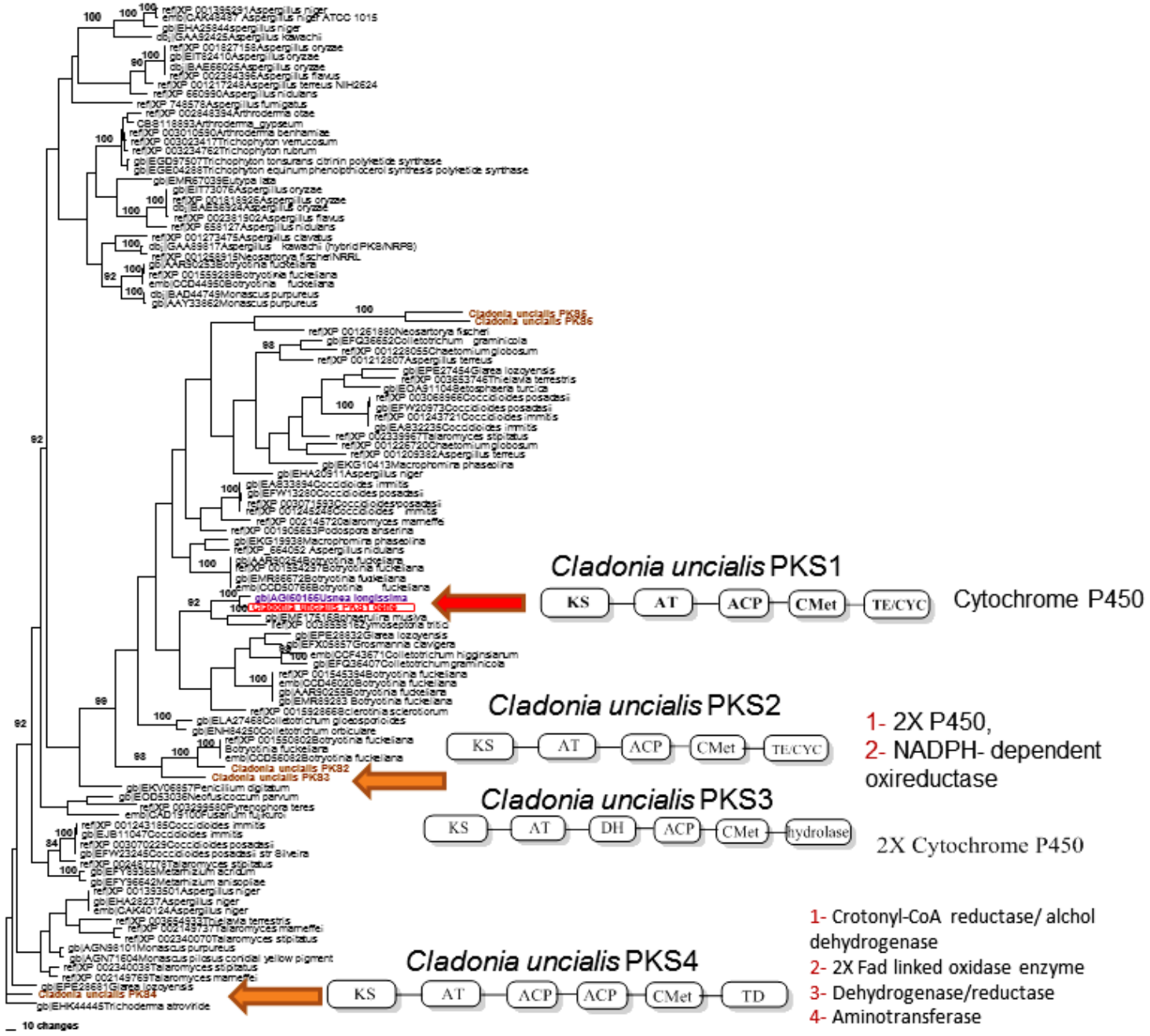


Figure 3.9 Phylogeny of the CYC domains from predicted *Cladonia uncialis* usnic acid PKS gene and homologous fungal non-reducing type I PKSs. Please refer to Appendix A for a high resolution one.

3.7 Conclusion

The gene cluster that is responsible for the biosynthesis of usnic acid in *Cladonia uncialis* was identified. In this chapter 32 PKS genes were identified; however, the lack of a required CMet domain helped to exclude all but five genes. Of these five remaining genes, only one gene was associated with an oxidative gene, consistent with the prediction that a single oxidative enzyme is responsible for dimerization of two molecules of MPA into usnic acid. This led us to suggest that this gene cluster codes for the expression of the protein products, MPAS and MPAO, comprising the complete biosynthetic machinery required for biosynthesis of usnic acid. Transcription of both *mpas* and *mpao* were demonstrated in an isolate of *C. uncialis* where usnic acid was the only detected metabolite. The phylogenetic analysis of two separate domains in *mpas* shows that this gene groups closely with other NR-PKS genes, and particularly, of those PKSs belonging to a known usnic acid-producing species of lichen. More usefully, *mpas* might serve as a probe for other similar genes in lichens and other fungi. This is the first report of the linking of a gene cluster to metabolite production in lichen and the first identification of the gene cluster responsible for usnic acid biosynthesis in lichen. This work will provide the knowledge required for the functional heterologous expression of these genes.

4 Chapter 4

Sequencing of a Gene Cluster that Codes for a Halogenated Polyketide from the Lichen fungus, *Cladonia uncialis*

4.1 Introduction

Halogenated natural products have long been a subject of research because of their potent and diverse biological activities. Approximately 5000 halogenated natural products have been discovered, with a structural complexity and diversity that rgants this group of compounds useful bioactivities for medicine and agriculture. (Wagner, C. *et al.*, 2009). Enzymatic halogenation is also proving to be an efficient alternative to classic synthetic approaches for creating organo-halides (Blasiak, C. *et al*., 2009, Neumann, S. *et al*., 2008, Murphy, D., 2006). An enzymatic approach to the synthesis of halogenated analogues of naturally occurring natural products can resulted in improved bioactivities (Eustáaquio, S. *et al*., 2003). Structural classes for these molecules range from simple phenolic and aliphatic compounds to complex polyketides and oligopeptides. A better mechanistic under-

standing of how these compounds are biosynthesized will not only further our understanding of the natural biological halogenation process but also give us a chance to develop alternative methods for the synthesis of halogenated analogues as biologically active metabolites.

In this chapter, the sequence and organization of a complete gene cluster coding for a halogenated natural product is reported. Whole gene sequencing from cDNA is also reported by using a technique called Rapid Amplification of cDNA ends (RACE), which provides an inexpensive and powerful tool to quickly obtain full-length cDNA when the sequence is only partially known. Starting with an mRNA mixture, gene-specific primers generated from the known regions of the gene and a non-specific anchor allows sequences to be identified in a short time. This method was first reported by Frohman *et al*., (1988) and has now become the most popular and effective method in terms of time and cost than cloning from plasmid or phage libraries (Schaefer 1995).

4.2 Materials and Methods

Taxonomic identification of *Cladonia uncialis* is given in Chapter (3.2.1)

Whole genome sequencing of *C. uncialis* mycobiont is given in Chapter (3.2.2)

4.2.1 Total RNA extraction and cDNA synthesis

The dry lichen sample was rehydrated with distilled water and incubated for two hours in natural light at room temperature. Total RNA was extracted from whole lichen thallus us-

ing the RNeasy Plant Mini kit (Qiagen, Mississauga, ON, Canada) following the manufacturer's directions. Approximately 100 mg of fresh thallus was lysed by freeze-grinding with liquid nitrogen. To degrade DNA, one microgram of RNA was treated with DNase I (Invitrogen, Burlington, ON, Canada) according tothe manufacturer's instructions. The synthesis of cDNA from total mRNA was done by using oligo dT primers and the ThermoScript RT-PCR System (Invitrogen, Burlington, ON, Canada) following the manufacturer's instructions.

4.2.2 Isolation of genomic DNA, amplification, and sequencing

Total genomic DNA was extracted from finely ground lichen material using a CTAB (Cetyl Trimethyl Ammonium Bromide) DNA protocol (Cubero et al. 1999) and the DNA was resuspended in 50 µL of warm sterile distilled water and stored at -20°C until required for experimental work.

Amplification of PKS gene fragments from *Cladonia uncialis* was performed using 1X phusion HF buffer (New England Biolab), 200 µM of each dNTP (Promega) 0·5 µM of each primer, 2 Unit/ µL of Phusion DNA Polymerase (New England Biolab), 3% DMSO, about 50 ng of total DNA, and water to 50 µL. The PCR conditions were denaturation at 98 ºC for 1 minute, then 30 cycles of denaturation 98 ºC for 10 seconds, annealing at 52 ºC-56 ºC for 30 seconds, and extension at 72 ºC for 1 minute. The reaction ended with a longer final extension at 72 ºC for 5 minutes. The annealing temperature varied according to the primers used in the PCR reaction. Three pairs of primers were designed to amplify part of the KS domain. A fourth primer pair was designed to amplify the gene from the 5'

end to the beginning of the AT domain. The primers PKS1F and PKS1R (Fontaine *et al.* 2010) were designed to amplify a single product (approximately 600 bp.) from the KS domain. The primers, PKS-DA-13F & PKS-DA-13R (Timsina et al., 2014) were designed to amplify 350 bp. An additional pair of primers (PKS1F1 and JSAT2R) (Table 4.1) was designed in this study to amplify 2581 bp from the 5' end of the non-reducing PKS gene to the beginning of the AT domain. The forward primer PKS1F1 was designed to match the beginning of the ORF region from *Xanthoria elegans* PKS1 (ABG91136). The reverse primer JSAT2R was designed from the previously sequenced *C. uncialis* PKS fragment. The size, quantity, and purification of the amplified products were performed by gel electrophoresis. The bands were excised from the gel and column purified using the Wizard SV Gel and PCR Clean-up system (Promega, Fisher Scientific, Nepean, ON, Canada), following the manufacturer's protocol. To quantify the DNA, the intensities of the bands were compared with that of a known DNA quantity using a 1 Kb plus DNA ladder (Invitrogen, Burlington, ON, Canada). All the primers used in this study are listed in Table 4.1. Sequencing reactions of the purified PCR product were performed in a total volume of 20 µL, consisting of 50 ng of purified DNA and BigDye 3.1 (Applied Biosystems, Foster City, CA, USA) following the manufacturer's instructions. Post reaction cleaning of the sequencing reactions was performed in a 1.5 mL Eppendorf tube by adding 0.25 volumes of 125 mM EDTA and 3 volumes of pure ethanol to the sequenced product. The samples were left at room temperature for 15 minutes and then centrifuged for 45 minutes at 4000 rpm. The supernatant was decanted by pipette and 60 µL of cold 80% ethanol was added. Once again the tubes were centrifuged for 15 minutes at 4200

rpm, followed by removal of the supernatant and drying the DNA in a ThermoSavant DNA 120 Speed Vac Concentrator (GMI INC., Ramsey, Minnesota, USA). The dried DNA was resuspended in 20 µL of Hi-Di formamide (AB) and loaded into a 96-Well plate for sequencing with a 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

PRIMER NAME	**GENE**	**PRIMER SEQUENCE FROM 5' TO 3'**
LC3cF	KS domain of PKS	GCIGARCARATGGAYCCICA
LC5cR	KS domain of PKS	GTIGAIGTIGCRTGIGCYTC
PKS1F	KS domain of PKS	TACGAAGCCCTAGAAATGGC
PKS2R	KS domain of PKS	ACGTTTGGCAGTTTCCTGTC
PKS-DA-13F	AT domain of PKS	GACCTCCACCAGCTTTCAAT
PKS-DA-13R	AT domain of PKS	GTGCATCTCGACATAGCT
PKS1F1	KS domain of PKS	ATGATGCCGCCCAACATGAA TATC
JSAT2R	AT domain of PKS	GATCTCAGCTT-GGACTGATCTC

Table 4.1 Primer sequences used for amplification and sequencing of PKS gene fragments used in this study.

4.2.3 Extraction of Total RNA, cDNA synthesis and reverse transcription

The dry lichen sample was rehydrated with distilled water and incubated in natural light at room temperature for two hours before performing the RNA isolation. Total RNA was extracted from whole lichen thallus using the RNeasy Plant Mini kit (Qiagen, Mississauga, Ontario, Canada) following the manufacturer's directions. Fresh thallus (100 mg)

consisting of young tissue at the branch tips was first ground in liquid nitrogen and after the procedure was completed the high-quality RNA was eluted in RNase-free water. RNA was quantified on a NanoDrop 2000C (Thermo Scientific spectrophotometer, Wilmington, USA) and 1µg RNA was treated with DNase I (Invitrogen, Burlington, ON, Canada) following the manufacturer's instructions. The synthesis of cDNA from total mRNA was done by using oligo dT primers and the ThermoScript RT-PCR System (Invitrogen, Burlington, ON, Canada) following the manufacturer's instructions.

4.2.4 RACE-cDNA synthesis

RNA extracts were quantified via NanoDrop 2000 C (ThermoScientific, Wilmington, NC, USA). The 260/280 nm absorptivity ratio was 2.1, and the 260/230 nm ratio was 2, which were suitable for commencing with RACE (rapid amplification of cDNA ends; Yeku & Frohman, 2011). 3'RACE (Figure 4.1) and 5'RACE-cDNA synthesis (Figure 4.2) was performed on the total RNA extraction of the mycobiont using the FirstChoice RLM-RACE kit (Invitrogen, Burlington, ON, Canada) according to the manufacturer's protocols, using the 3' RACE primer for reverse transcription of total mRNA into cDNA. Primer sequences are listed in Table 4.1. Gene-specific primers KSATF (forward) and KSATR (reverse) (Table 4.1) were designed from the sequenced cDNA and used in conjunction with the complementary RACE primer to amplify the 3' and 5'-ends of the PKS cDNA. One µg of total RNA was used for reverse transcription by moloney murine leukemia virus (M-MLV) Reverse Transcriptase using the 3' RACE Adapter as in Table 4.1; product from the reverse transcription reaction (1 µL) was used for PCR along with 1 U

of fusion DNA polymerase (New England Biolab), one of the 3' RACE outer primers which is complimentary to the anchore sequence. The first RACE-PCR was used as a template for the nested PCR by using different genes specific primers as in Table 4.1 along with 3' RACE Primer (inner primer). Thirty three different reactions by nested PCR were performed to give the full 3' RACE length. All gene-specific reactions for amplification of the cDNA were performed in duplicate. The total volume of the reaction was 50 µL. Almost 80 ng of cDNA was added to 1X long amplification polymerase buffer, 200 µM of each dNTP, 0.5 µM of each primer, 2 Unit/ µL long amplification polymerase (New England Biolab, Canada) and water up to 50 µL. The PCR condition for the 3' RACE –PCR reactions was as follows: initial denaturation at 95 ^{o}C for 2 minutes followed by 30 cycles of 94 ^{o}C for 30 s, 15 s at (56 ^{o}C to 60 ^{o}C) as annealing temperature depending on the primer used and then 65 ^{o}C for 4 minutes as extension. The final single cycle was a final extension at 65 ^{o}C for 10 minutes.

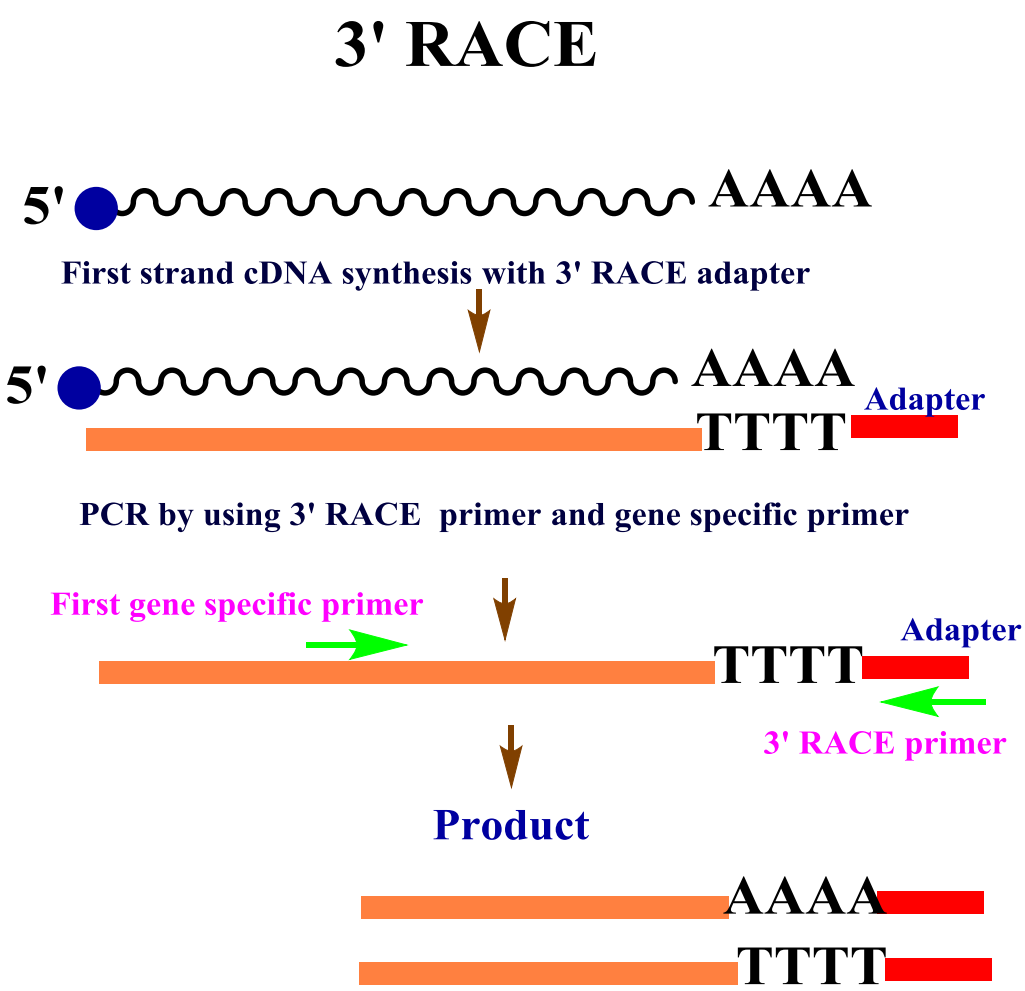


Figure 4.1 The identification of the 3' end of an unkown sequencing of cDNA using RACE technique. 3´ RACE - cDNA synthesis Poly (A) tail region was used as an initial priming site. Gene specific primer was used with the 3' RACE to amplify the desired product.

The 5' RACE, cDNA synthesis (Figure 2.5) was initiated by dephosphorylation of the mRNA by using calf intestinal phosphatase (CIP). This step is essential to dephosphorylate degraded mRNAs, which are uncapped at their termini. The full-length mRNAs, which have methyl-G caps at their termini, are not affected by the dephosphorylation process (Volloch, V., *et al.*, 1991). This makes sure that the degraded RNA does not contribute in the ligation step, because the phosphate group is required to drive the reaction.

The full-length mRNAs are then decapped using tobacco acid pyrophosphatase (TAP), which leaves them, with an active and phosphorylated 5' terminus (Mandl, C.W. *et al*., 1991). Using T4 RNA ligase, this mRNA is then ligated to a 45 base pair RNA adapter. Reagents and conditions for 5' RACE were the same as the 3' RACE except the extension time was 2 minutes and 30 seconds instead of 4 minutes. After sequencing the 3'RACE and the 5'RACE products, a series of internal primers were designed from the sequences in Table 4.1. A gene walking approach was applied to amplify the whole gene sequence from the cDNA. The PCR conditions for the gene walking PCRs were varied depending on the primers and the enzymes used in each reaction.

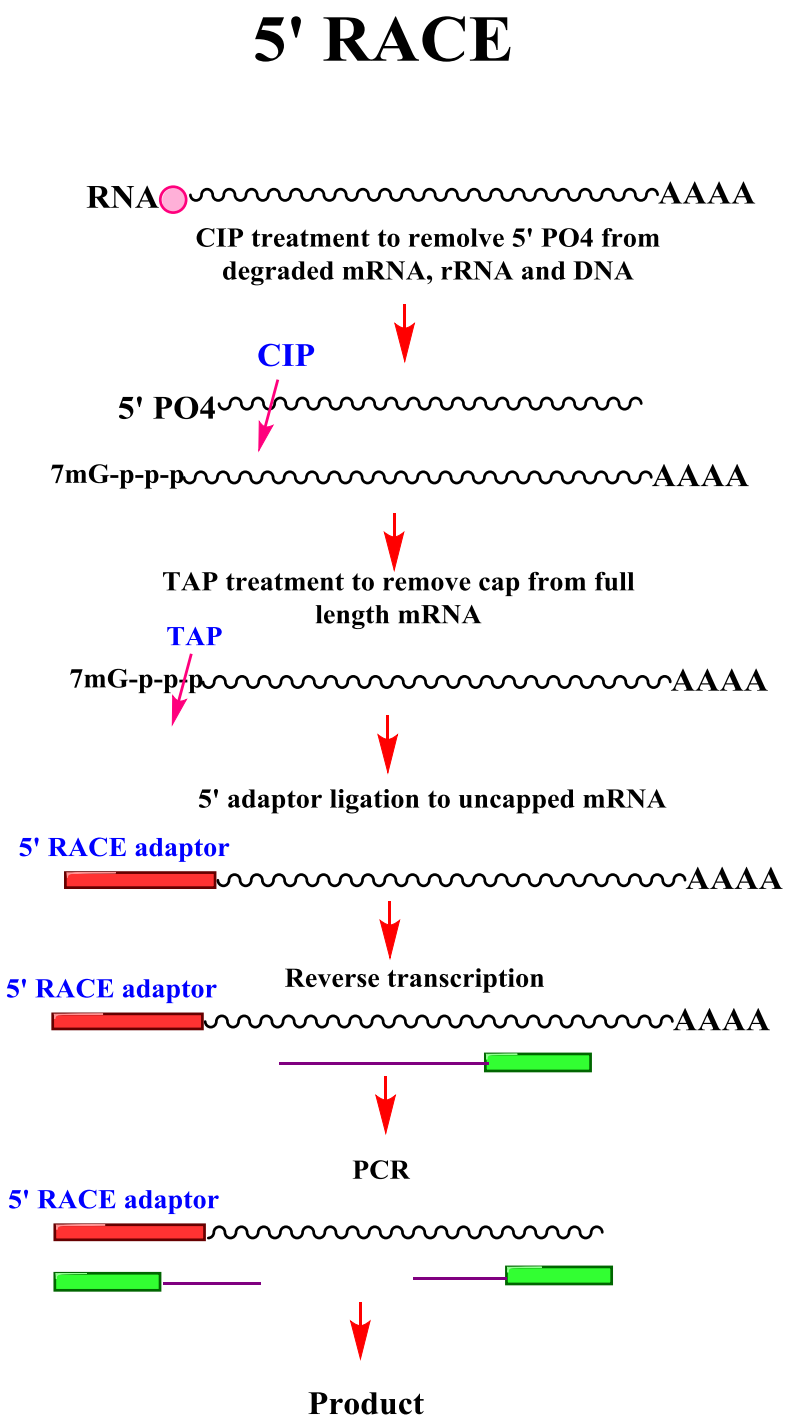


Figure 4.2 5′ RACE - cDNA synthesis started by dephosphorylation of the mRNA by using calf intestinal phosphatase (CIP) followed by mRNA decapping using tobacco acid pyrophosphatase (TAP) followed by tagging of the 5' end by ligation of the 5′ RACE adapter and then cDNA synthesis.

4.2.5 Cloning and sequencing of the RACE products

The 3’ RACE and 5’ RACE products were ligated into pGEM T easy cloning vector (Chapter 2, 2.13.2) by using 1X rapid ligation buffer, T4 DNA ligase, 50 ng pGEM T easy cloning vector, 1 µL T4 DNA ligase, 200 ng PCR product to bring the total reaction to 10 µL. The ligation reaction was kept at 4^{o}C for 4 hours and transformed into JM109 high efficiency competent cells (Fisher Scientific, Canada).

Transformation with the 3’RACE and 5’ RACE colony ligation products yielded a high number of colonies. The DNA sequences were confirmed at each step by PCR amplification, purification and sequencing. Plasmid purification was carried out using the Wizard_plus SV Miniprep system (Promega) by strictly following the manufacturer’s instructions. The purified plasmids were then sequenced by the Genewiz DNA sequencing company in the USA. The resulting nucleotide sequences were compared with each other to assemble a complete PKS gene sequence using Sequencher v. 4.8 (Gene Codes Corporation, Ann Arbor, MI, USA).

PRIMER NAME	PRIMER SEQUENCE FROM 5’ TO 3’ ENDS
Sample 1-RACE beg_1R	GTTGACGTCCGTCTCTGTCA
Sample 1RACE beg_1R_R	ACCAAGATATCGCCCACAAC
KS1F1-RACER2-M13R[1]	CGGACACGTCTCTTCCAGAT
Sample 1RACE beg_1R_R-1	GTTGTGGGCGATATCTTGGT
KS1F1-RACER2-PKS1F1-MR	GGGCATGAAACCTGACATCT
2010-10-14-H09-RACER2F1	AAGCCTACCAAACCGATCCT

2010-10-14-A10-RACER2F2	TATGGCCAGACGAGTGATGA
PKS1F1-RACE-R2-M13F(21)[1]	AAGGTGCCAACTCGGTCTAA
2010-09-E05-PKS1F(RACE_R)	GTCGATGTTTTGTGCTGCAT
2010-09-03-G09_Mona_10_F	TCGATGTTTTGTGCTGCATT
2011-09-21_Bo3	CATTCCCGGGACTATACGTG
2011-09-21_F02	TATGGCCAGACGAGTGATGA
F2-F3-5-3RACEF3	TTAGCCTGCACATCTCTCCA
F2-F3-16-3RACEF3	ATACTGCGGACGGTTTCTGT
2011-03-03_C02_M3-3RACEF3	CCTGCAGGACATTTGACGAT
2011-09-21_C03	GGAGCTCTTGAGTTGGAGCA
2011-09-21_E02	CGTCGACACCAGTGTTTCTC
F2-F3-16-3RACE F16(4)	ATCCAGCAATACCCGTCATC
F2-F3-16-inner	GGCGAATCGATCAAGAAGAG
F2-F3-16-RACE_F_17_R	AAGAGCATATCTCGGGACGA
F2-F3-5-inner	AATTGAGGCCTCGAAATGTG
F2-F3-9-RACE_end_2	CTCGTGAATGGCATCTCTGG
SAMPLE 8-RACE-END	CTGCTCTCTCCCTCGTGAAT
RACE_end_2F_inner-B-RACE inner	GGCGAATCGATCAAGAAGAG
RACE_end_2F_inner-D-RACE_end	TGCTCTCTCCCTCGTGAATG
RACE_end_2F_inner-E-RACE_end_2F	GCAAGTGGCTGTAATGTGTCTG
RACE_end_2F_inner-E-RACE_end_2F_R	AGAGCAAGTGGCTGTAATGTG
RACE_end_2F_inner-B-RACE_end_2F	CGCGATGGACTGTGTAAAGA
RACE_end_2F_inner_ist-	TGCTCTCTCTCCCTCGTGAATG

RACE_end_2F	
RACE_end_2F_inner-C-RACE_end_2F	TCTGGAGATGCAGAGGTAAGG
2011-11-21_A02	GCAAGTGGCTGTAATGTGTCTG
RACE_end_2F_inner_2nd-RACE_end_2F	CGTTCCCCATTGTAAACCTC
2011-06-14_C06_RACE_END_2F	GAGCAAGTGGCTGTAATGTGTC
RACE-END2FC	TGACGGAGATGACATGGTTG
5RACE-OUTER	ACAAACCTCCGACTCATCG
3RACE-INNER	CGCGGATCCGAATTAATACGACTCACTATAGG
5RACE-INNER	CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG

Table 4.2 Primers employed in gene walking expermintal work.

4.3 Results and Discussion

Four pairs of primers were used to amplify different PKS gene fragments from *Cladonia uncialis* genomic DNA. The first three pairs were designated to target the KS domain while the fourth pair of primers (PKS1F1 and JSAT2R) were designed to target the beginning of the KS to the middle of the AT domain (Table 4.1). The PCR amplification of *Cladonia uncialis* genomic DNA by using these four pairs of primers yielded four bands, a 2581 bp band labeled PKSI, a 644 bp band labeled PKSII, 350 bp band labeled PKSIII and 747 bp band labeled PKSIV. The first three bands PKSI, PKSII, PKSIII were identified by using homology analysis BLAST search as nonreducing PKS genes. The fourth band PKSIV is identified as a partly reducing PKS gene.

In order to test the transcription for each of the four PKS genes, the cDNA library from RT-PCR was used as a PCR template and the PCR reactions were done using the four sets of primers used to amplify the genomic DNA (Table 4.1). Only the cDNA corre-

sponding to the PKSI gene was amplified and no other products could be detected from the three other genes. To get the full PKSI gene, the RACE-cDNA synthesis technique was used to amplify the full length. Primers for each RACE cDNA synthesis are listed in Table 4.2. Gene-specific primers (GSP) were designed for the RACE PCR starting at the known KS domain sequence. RACE products were obtained using gene specific primers of the KS domain of the targeted PKS. Nested PCR yielded a 5' RACE product of almost 2800 bp, which was subsequently cloned. The PCR for the 3'RACE resulted in a main product at 4 kb. This band was excised from the gel, cloned and sequenced by using nested PCR. All the primers are listed in Table 4.2. The alignment and overlapping of all the PKSI cDNA partial sequences resulted in a sequence of 6000 bp in total (Figure 4.3). The sequence alignment and assembly was done using the Sequencher v. 4.8 program. The Sequencher v. 4.8 was adjusted to align and assemble the sequences that match each other by 90% which allow only the high quality sequence to overlap with each other. The first sequence to be aligned in the PKSI gene was 724 bp length (Sample 1-RACE beg_1R) that was located at the 5'end of the gene (Figure 4.3). The alignment continued with the second sequence (KS1F1-RACER2-M13R [1]) which increased the PKSI cDNA fragment length from 724 bp to 1500 bp. The third sequence added was (KS1F1-RACER2-PKSIF1-MR) which increases the length to 1700 bp. The gene length increased to 2600 bp by adding the third sequence (PKSIF1-RACE-R2-M13F (21)[1]). Another sequence (F2-F3-16-3RACEF3) was added by the Sequencher to bring the total length into 3300bp. The gene length increased to be 4400 bp when the sequence (F2-F3-16-3RACE F16(4)) was aligned to the previous sequences. The PKSI gene length was 4800 bp after

adding (F2-F3-16-inner) sequence. Two cDNA sequences (F2-F3-5-inner and RACE_end_2F_inner_ist-RACE) were added by the Sequencher program to bring the total PKSI cDNA length to 6148 bp (Figure 4.3).

The National Center for Biotechnology Information (NCBI) Open Reading frame (ORF) finder was used to identify the ORF and the translated amino acid sequence was used for alignment with the conserved domain database (CDD). Several conserved protein domains are identified and the predicted protein was classified as a type 1 fungal non-reducing PKS (NR-PKS) enzyme. The PKSI cDNA sequence was analysed by Antibiotics and Secondary Metabolites Analysis Shell V. 2.0 (antiSMASH) program designed to detect and annotate secondary metabolite biosynthetic gene clusters (Medema, H. et al., 2011, Blin, K. et al., 2013). The gene sequenced was confirmed to be a type I iterative non-reducing polyketide synthase containing all essential domain components necessary for iterative condensation as well as a terminal thioesterase domain (Figure 4.4). Annotation by the program yielded two acyl carrier protein (ACP) domains. Recent research in polyketide assembly has identified a product template (PT) domain that is now known to be responsible for controlling the number of iterative condensations as well as the timing and folding of the polyketide chain for aldol cyclizations (Crawford et al., 2009; Crawford et al., 2008; Crawford & Townsend, 2010). As research is in its infancy, antiSMASH is ill-programmed to recognize most PT domains, and for this reason the annotation of the PT domain done manually by aligning the PKS gene sequence with other known product template domain sequences from Aflatoxin biosynthesis polyketide syn-

thase (Crawford, M. et al., 2009). No reduction domains were detected within the PKS gene.

The characterization of the biosynthetic gene cluster was achieved by *de novo* whole-genome sequencing of the mycobiont of *C. uncialis*. The assembly of short reads into contigs that could be interpreted by gene annotation programs is described in Chapter 3. By searching the whole-genome transcripts for a nucleotide sequence that matches the one obtained from the cDNA, the contig that contained the gene could be identified using the program antiSMASH. This search of the genome also resulted in the presence of several tailoring genes adjacent to the PKS gene. These accessory genes were an upstream FAD-binding mono-oxygenase as well as a downstream halogenase and O-methyltransferase. A sequence of DNA, annotated as either dehydratase-like or ketoreductase-like in function, was found upstream of the PKS. Comparison of the genomic DNA sequence with the cDNA sequence revealed a single intron 46 bp in length located between the KS and AT domains.

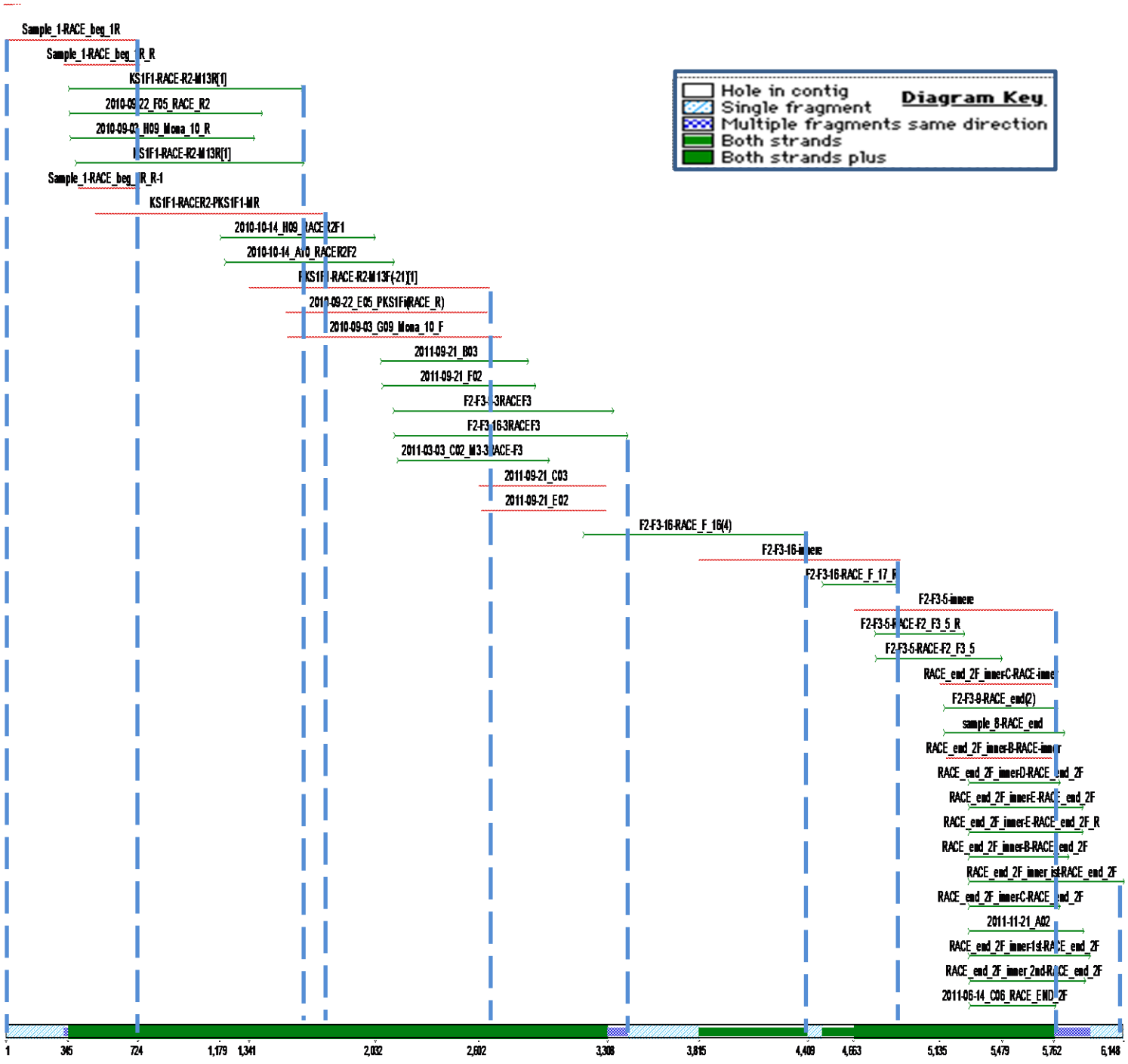


Figure 4.3 PKSI cDNA sequence fragment alignment by using Sequencher v. 4.8. The green lines represent the cDNA fragments that were sequenced by using forward primer sequences while the red lines represent cDNA fragments that were sequenced by using the reverse primer.

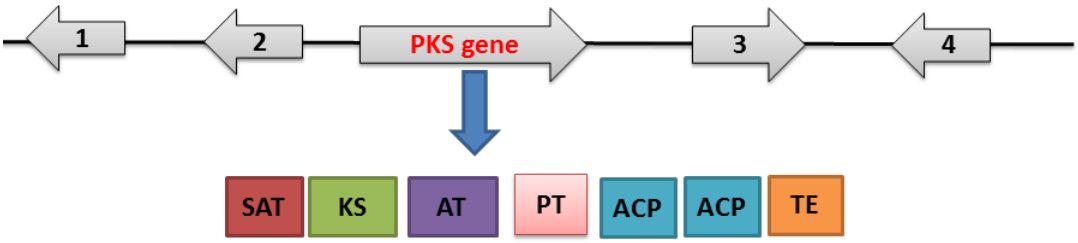


Figure 4.4 The Cladonia uncialis halogenase-associated polyketide synthase gene cluster responsible for the biosynthesis of an unidentified halogenated secondary metabolite. (1) FAD-binding monooxygenase; (2) Ambiguous dehydratase-like or ketoreductase-like gene; then polyketide synthase, containing starter acyltransferase (SAT) ketosynthase (KS), acyltransferase (AT), (PT) product template, two acyl carrier proteins (ACP), and terminal thioesterase (TE) domains; (3) halogenase; (4) O-methyltransferase.

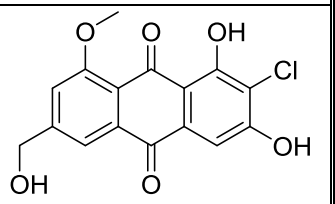
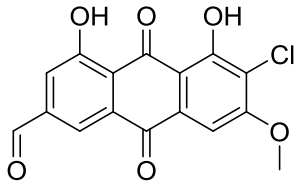
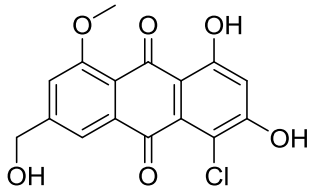
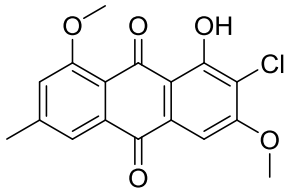
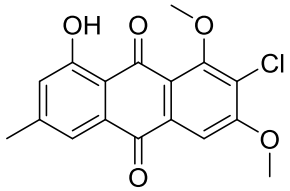
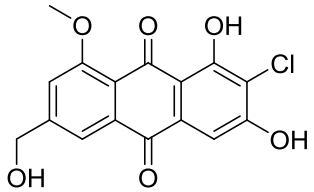
The HPLC analysis of the acetone extract from the field-collected *Cladonia uncialis* displays a single secondary metabolite, the non-halogenated dibenzofuran usnic acid (Chapter 3). Nonetheless, it was possible to detect the mRNA associated with the newly identified PKS by reverse transcriptase amplification of the cDNA. This implies that the PKS gene is transcriptionally functional, but not translated into active enzyme or that the lichen does produce the metabolite but is below the HPLC detection limit or the metabolite couldn't be extracted by the acetone extraction. It may be possible that this gene is functional but only up-regulated by particular environmental or chemical stimuli (Zazopoulos, E. et al., 2003). Considering the importance of halogenated natural products and the additional fact that to the best of best knowledge no PKS gene has been linked with a haloginated metabolite in lichen, I decided to attempt to identify plausible lichen metabolites that would be consistent with the biosynthetic operations of the PKS and surrounding genes. The Dictionary of Natural Products (DNP) is a comprehensive database of the

more than 250,000 natural products identified from all forms of life to date and was used to search for lichen halogenated polyketides that could be used as candidate structure types for the unidentified polyketide. From the 'Add Property' list, the search restriction for 'Type of Organism' to present natural products discovered from lichen (Code ZK0001) was included. An additional search restriction was applied by having the DNP report only those molecules that contain at least one fluorine, chlorine, bromine, or iodine atom, using the 'molecular formula by element' search function. The search of polyketide natural products was not restricted because the aim was to maximize search returns in the event that the compounds were not catalogued as polyketides upon entry into DNP due to ambiguous metabolic origins. This search yielded 0 fluorinated, 116 chlorinated, 12 brominated, and 0 iodinated lichen products. The PKS gene possesses the minimum domain components (KS, AT, ACP) with no reductive domains (none of KR, DH, and ER). The gene cluster does possess an upstream FAD-binding monooxygenase as well as a downstream O-methyltransferase and halogenase. Hence, from this 128 total candidate natural products, the list was narrowed by selecting only those metabolites that: (1) were known polyketides or had structures reminiscent of polyketide cyclization; (2) did not require reduction during chain elongation; (3) were O-methylated; (4) were oxidized by addition of a hydroxyl group; and (5) did not require catalytic operations inconsistent with the gene cluster, for example, oxidative dimerization or esterification.

Only one type of lichen halogenated polyketide, chloro-O-methyl-anthraquinones, was consistent with all biosynthetic criteria (Table 4.3). Examples of other lichen polyketide organohalides listed in the DNP, and their reason for exclusion, are listed in (Table 4.4).

Chlorinated anthraquinones are commonly found in lichens as pigments, producing many of the brilliant colours found in nature (Mathey *et al.*, 2002, Søchting, U., 1997). Anthraquinones protect lichens from UV radiation, and the biosynthesis of these metabolites have been observed to increase in response to UV radiation (Gauslaa, Y. et al., 2005, Solhaug, A. et al., 2003). It seems reasonable that a PKS producing an anthraquinone derivative would be encoded by *Cladonia uncialis* and that this gene would be induced by stresses encountered in the natural habitat. Confirmation of this prediction and the functionality of this PKS would require expression of the PKS followed by successful detection and characterization of the metabolite, or, functional heterologous expression of the gene cluster and isolation of the metabolite in fermentation culture.

The hypothesized identity of the metabolite is further supported by known folding mechanisms of anthraquinones and the decarboxylative ring closures common to terminal thioesterase domains. Eight condensation reactions between an acetyl-CoA primer and malonyl-CoA extension units produces a 16-carbon polyketide that may then be folded in at least two ways to produce anthraquinones (Figure 4.5). Terminal thioesterase domains within PKSs perform the same decarboxylation reaction that would be necessary for the final ring closure in anthraquinone biosynthesis (Gessler, N. et al., 2013). Hence, the hypothesis of an anthraquinone product is consistent with the domain organization of the PKS with respect to both lack of reducing elements as well as type of terminal domain.

IUPAC/COMMON NAME	LICHEN(S) OF ORIGIN	STRUCTURE
2-chloro-1,3-dihydroxy-6-(hydroxymethyl)-8-methoxyanthraquinone	*Nephroma laevigatum*	
6-chloro-9,10-dihydro-4,5-dihydroxy-7-methoxy-9,10-dioxo-2-anthracenecarboxaldehyde	*Cetraria sp.*	
1-chloro-2,4-dihydroxy-7-(hydroxymethyl)-5-methoxyanthraquinone	*Nephroma laevigatum*	
2-chloro-1-hydroxy-3,8-dimethoxy-6-methylanthraquinone	*Nephroma laevigatum, Astroplaca opaca*	
2-chloro-8-hydroxy-1,3-dimethoxy-6-methylanthraquinone	*Caloplaca xanthaspidis*	
2-chloro-1,3-dihydroxy-6-(hydroxymethyl)-8-methoxyanthraquinone	*Nephroma laevigatum*	

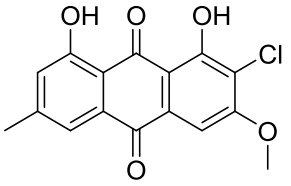
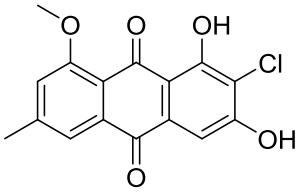
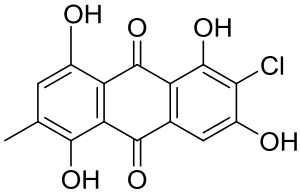
2-chloro-1,8-dihydroxy-3-methoxy-6-methylanthraquinone	*Sphaerophorus fragilis, Sphaerophorus globosus, Nephroma laevigatum, Byssoloma tricholomum,Calopla caarenaria,*	
2-chloro-1,3-dihydroxy-8-methoxy-6-methylanthraquinone	*Nephroma laevigatum, Fusarium aquaeductuum*	
2-Chloro-1,3,5,8-tetrahydroxy-6-methylanthraquinone	*Lasallia papulosa*	

Table 4.3 Sample lichen halogenated polyketides cataloged in the Dictionary of Natural Products database that are consistent with the biosynthetic logic of the PKS gene cluster.

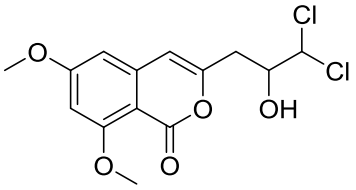
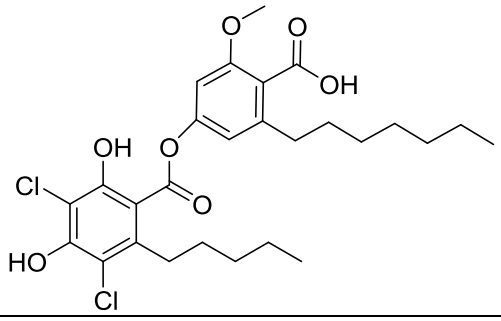
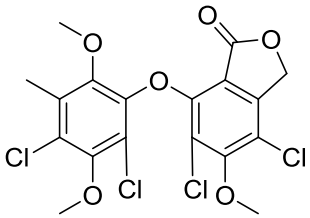
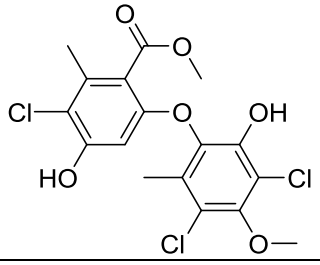
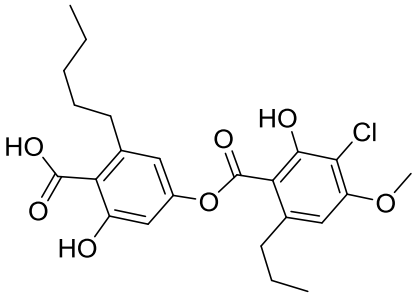
IUPAC/COMMON NAME	STRUCTURE	REASON FOR EXCLUSION
3',3'-Dichloro-8-O-methyldiaporthin		No monooxygenase modification
3,5-Dichloro-2'-O-methylnorstenosporic acid		Requires oxidative dimerization, no monooxygenase modification
Canesolide		Requires C-methylation, requires oxidative dimerization
Buellin		Requires C-methylation, requires oxidative dimerization
3-Chlorostenosporic acid.		Requires oxidative dimerization

Table 4.4 : Sample lichen halogenated polyketides cataloged in the Dictionary of Natural Products database that are excluded as candidate gene products.

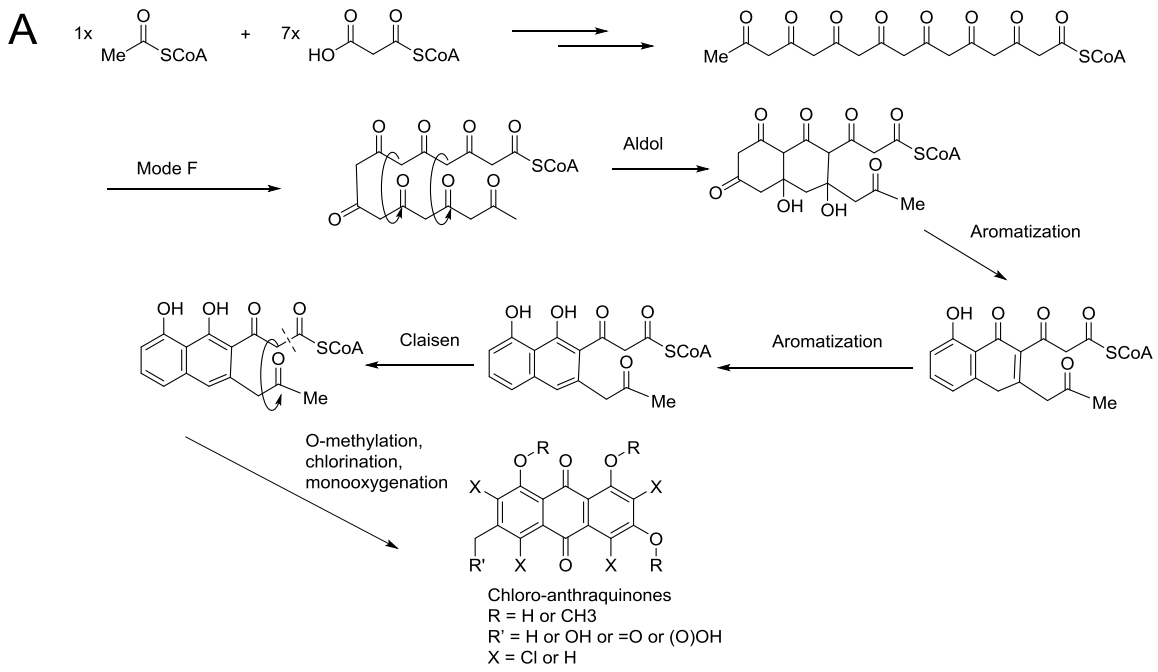


Figure 4.5 Biosynthesis of anthraquinones proceeds through variant cyclizations of 16-carbon polyketides. Formation of the final ring proceeds through decarboxylation (Gesslera et al., 2013, Bringmann et al., 2009).

4.4 Conclusion

In this Chapter for the first time, a complete sequence of a polyketide synthase (PKS) gene and accompanying accessory genes from the lichen, *Cladonia uncialis* has been reported by *de novo* whole-genome DNA sequencing of the mycobiont as well as cDNA sequencing of the reverse transcribed mRNA via rapid amplification of cDNA ends (RACE). BLAST and antiSMASH analyses revealed that the PKS is composed of only the minimal domain components (ketosynthase, acyltransferase, acyl carrier protein), but is flanked by a downstream halogenase and O-methyltransferase as well as an upstream monooxygenase and a reductive gene of indiscernible function. Successful reverse tran-

scription of the mRNA into cDNA demonstrates that the PKS is transcriptionally active; however, no halogenated secondary metabolites have been detected in *C. uncialis* extracts. Based on the biosynthetic operations of the gene cluster as well as catalogued examples of halogenated polyketides, this suggests that the gene cluster is transcriptionally active that could be responsible for anthraquinone biosynthesis. Confirmation of the assigned metabolite will require functional heterologous expression of the gene cluster with an inducible promoter and production of the metabolite through a host fermentation culture.

5 Chapter 5

Effect of Culture Conditions on Secondary Metabolite Production in *Cladonia uncialis*

5.1 Introduction

Lichens, photosynthetic cyanobacteria or algae and fungal symbionts, have a long history in medicine for their use in the treatment of ailments as well as other traditional uses in cosmetics, perfumes, aphrodisiacs, and food (Shukla *et al*., 2010). It is known today that such properties are conferred to lichens through the presence of structurally diverse secondary metabolites, of which more than one thousand have been identified to date (Molnar *et al*., 2010). These metabolites have been purposed through nature to function as antifeedants, antioxidants, antibacterial compounds, among myriad other functions (Shukla et al., 2010; Boustie & Grube, 2007; Huneck, 1999; Muller, 2001). Lichens are remarkable for occupying desolate habitats despite adverse atmospheric conditions. Consequently, lichens have been adapted to grow slowly, in many cases less than 1 cm/year of growth is observed, limiting efforts to harvest lichen metabolites for human use. Tech-

niques such as applying different conditions in the mycobionts cultures may activate the silent genes and unlock the full potential of the lichen metabolome (Wasil *et al*., 2013).

The production of lichen substances in the culture mainly depends on culture conditions. Most experimental work indicates that these conditions are related to the composition of the nutrient medium, osmotic stress, temperature and exposure to high light intensities (Hager *et al*., 2008, Zocher *et al*., 2005 and Hamada *et al*., 1996). Investigations on polyketide synthase (PKS) genes have begun to develop a molecular approach to investigate these genes in lichens or in cultures. The ultimate goal of this work is to provide a sustainable source of lichen natural products to support the applications. The objective of the present investigation was to grow the lichen mycobiont of *Cladonia uncialis* under different culture conditions and examine secondary metabolite production.

Previous research has demonstrated that the presence of extra sucrose in the media supports the production of secondary metabolites (Hamada, 1993 and 1995). In this work, different sucrose concentrations were applied in different culture media and secondary metabolite production was examined. Trace elements media was also used in combination with different sucrose concentrations.

The secondary compounds were determined by using high performance liquid chromatography (HPLC).

5.2 Material and Methods

5.2.1 Lichen material and mycobiont isolation

The lichen-forming fungus used in this study was visually identified as *Cladonia uncialis* (Voucher number: Normore 8774, Herbarium, University of Manitoba, Winnipeg, Canada), collected in October 2008, from Northern Manitoba on a south-facing granite ridge (N54º 42' 24.7"; W101º 33' 53.1"). The mycobiont was cultured from ascospores using the method of Yoshimura *et al*., (2002) with modifications as described below: A single apothecum was removed from a thallus branch, washed under running tap water for 90 min, transferred to a 50 mL beaker and washed for 30 min with sterile distilled water using a magnetic stir bar. The wash was repeated in sterile distilled water by adding almost 150 µL of Tween 80. The water was changed and the wash was repeated three times. Finally apothecia were transferred to a sterile beaker containing sterilized distilled water and stirred using a magnetic stir bar for another 15 minutes. Single apothecia was then cut from the larger thallus on a glass slide using a sterile razor blade inside a laminar flow hood and affixed to the lid of an inverted petri dish with petroleum jelly to permit spore release onto 1.5 % (w/v) water-agar media. To ensure that the petri plate remained moist, which enhanced spore release, three to four triangular pieces of sterile wet blotting paper were placed between the apothecia on the petri plate cover. The culture plate was then sealed with parafilm, labelled, and incubated at 20 ^{o}C. The spores from each apothecium were transferred to freshly prepared malt agar medium (1.5 % malt and 1.4 % agar) using a dissecting microscope. A spore suspension was spread over the new media to ob-

tain single spore colonies by adding two to three drops of sterile distilled water to resuspend the spores. The colony growth was monitored over time on different media compositions for four months.

5.2.2 Fungal media

All media used in this chapter are listed in Table 5.1

COMPOUND NAME	GRAM/LITER
1.5% water agar media	
Agar	15g
Malt yeast extracts media (Ahmadjian, 1993)	
Malt extract	20g
Yeast extract	2 g
Agar	20 gram
Malt yeast extracts media with sucrose	
Malt extract	20g
Yeast extract	2 g
Agar	20 gram
Sucrose	10g, 15g,
Malt - sucrose agar media	
Agar	15 g
Malt extract	14 g
Sucrose	10 g
Malt agar media	
Agar	15 g
Malt extract	14 g
Trace element media (Gary A. Strobel *et al.*, 1974) (Stock solution)	
Compound name	Gram/10mL
Na_2SO_4	1.989
$FeCl_3$	0.051
$ZnSO_4$	0.0156
H_3BO_3	0.0136
NaI	0.00675
$CaCl_2$	2.63
KNO_3	0.799
$MgSO_4$	3.611
KCl	0.649
NaH_2PO_4	0.193

Table 5.1 Different media components that were used for the *Cladonia uncialis* mycobiont cultures.

In the trace element medium, 10 mL of each stock solution was added to 940 mL of distilled water to bring the final volume 1 Liter. Fifteen grams of agar were added to the one liter solution. The medium was sterilized at 121 ^{o}C for 20 minutes and then was left to cool at room temperature in the laminar flow hood and then poured in a petri dishes. The same medium was prepared with 15g/L sucrose.

5.3 DNA extraction and sequencing

To confirm the identity of the cultured mycobiont a fresh colony was placed in 1.5 mL microcentrifuge tubes and total DNA was extracted from the cultured mycobiont using Gen JET Genomic DNA purification Kit (Thermoscientific, USA) following the manufacturer's protocol. Confirmation of the identity of the mycobiont taxonomic classification was performed by amplification of the Internal Transcribed Spacer 1 (ITS1) of the nuclear ribosomal DNA (rDNA) by using the fungal primers1780F-5' forward (5'CTGCGGAAGGATCATTAATGAG3') (Piercey-Normore and DePriest, 2001) which anneal to the 18S (ribosomal small subunit) and ITS4-3' reverse (5'TCCTCCGCTTATTGATATGC3') (White *et al*., 1990) (Chapter 2, section 2.8). The ITS fragment was sequenced by MICB (Manitoba Institute of Cell Biology, Cancer Care Manitoba, University of Manitoba) DNA sequencing services.

5.4 High Performance Liquid Chromatography (HPLC) analysis of lichen cultures

5.4.1 Extraction of colonies

Mycobiont colonies grown on solid media, with different compositions, were extracted with acetone at room temperature for 20 minutes. After evaporation of acetone under vacuum, the extracts were suspended again in acetone to bring the final concentration to 1mg/mL. The same process was carried out on each different media alone as positive controls.

5.4.2 HPLC analysis of the extracts

HPLC analysis was performed on acetone extracts from the natural lichen thallus and from the cultured mycobionts. The HPLC system is the same as described in the general materials and methods (Chapter 2, 2.3). The first gradient was held at 20 % methanol to 80 % water in 0.075 % aqueous trifluoracetic acid for 10 minutes then a linear gradient to 80 % methanol and held at that composition for 20 minutes followed by a linear gradient back to 20 % methanol for 10 minutes and held there for more 10 minutes. The total run time was 60 minutes.The number of secondary metabolites within each sample was compared with those from the media alone, and were counted for each trace.

5.5 Results and discussion

The identity of the mycobiont cultures were confirmed by amplification and sequencing of the Internal Transcribed Spacer 1 (ITS1) of the nuclear ribosomal DNA (rDNA) which is widely used in taxonomy and molecular phylogeny because it has a high degree of variation even between closely related species. The nucleotide sequences were 100% identical with the nucleotide sequence obtained from the previously collected thallus of *Cladonia uncialis* and reported in Chapter 2, section 2.7.

The influence of various culture conditions on the production of different secondary metabolites was investigated including usnic acid of a lichen species *Cladonia uncialis* under laboratory conditions. Interestingly, it has been shown that all the cultured mycobiont mycelia grown in different media assayed, produced a variety of different secondary metabolites but not usnic acid (Fig. 5.2). There were more secondary metabolites detected in the acetone extract from the cultured mycobiont when compared to the acetone extract of the natural lichen thallus (Figure 3.2). There were at least 10 unidentified metabolites detected from the 8 month old cultured mycobiont by using HPLC. The appearance of different secondary metabolites in different lichen cultures under different culture conditions is consistent with the presence of a large number of gene clusters that were detected in the *Cladonia uncialis* genome sequence (Chapter 6). It appeared that some of the *Cladonia uncialis* gene clusters were activated under different culture conditions. The effect of different culture conditions, reflected by the production of secondary metabolites has been studied by Brunauer *et al*., (2007). They found that the number of the secondary metabolites increased in the cultured mycobiont when compared to the natural lichen,

Xanthoria elegans. Different studies also indicated that through the evolution of the microorganism, they developed the capacity to produce a large number of secondary metabolites through different pathways that are not expressed during the normal growth. The compounds they produce are largely considered to be nonessential to the organism. However, under stressful conditions, these pathways are induced to synthesize a variety of metabolites that are often required for the organism to survive (Brakhage, A.; 2011, Schroeckh, V., 2011). Our results are consistent with this research since a number of secondary metabolites were detected in the fungal culture that were not detected in the natural lichen.

More than 9 different unidentified secondary metabolites were detected by HPLC in the crude extract from the cultured mycobiont *Cladonia uncialis* that was grown in water and agar medium (Figure 5.1). The retention times of these compounds were 6.8; 22.2; 22.4; 23.4; 24.6; 25.0; 26.1; 28 7 and 30.0 minutes. Each lichen extract was standardized to a concentration of 1mg/mL. The low absorbance numbers represented by low peak values in Figure 5.1 indicate that the metabolites are present in low concentrations.

A number of crude extracts from different lichen cultures were subjected to NMR analysis for structural determination. However the spectra displayed signals consistent with a mixture of metabolites and could not be resolved.

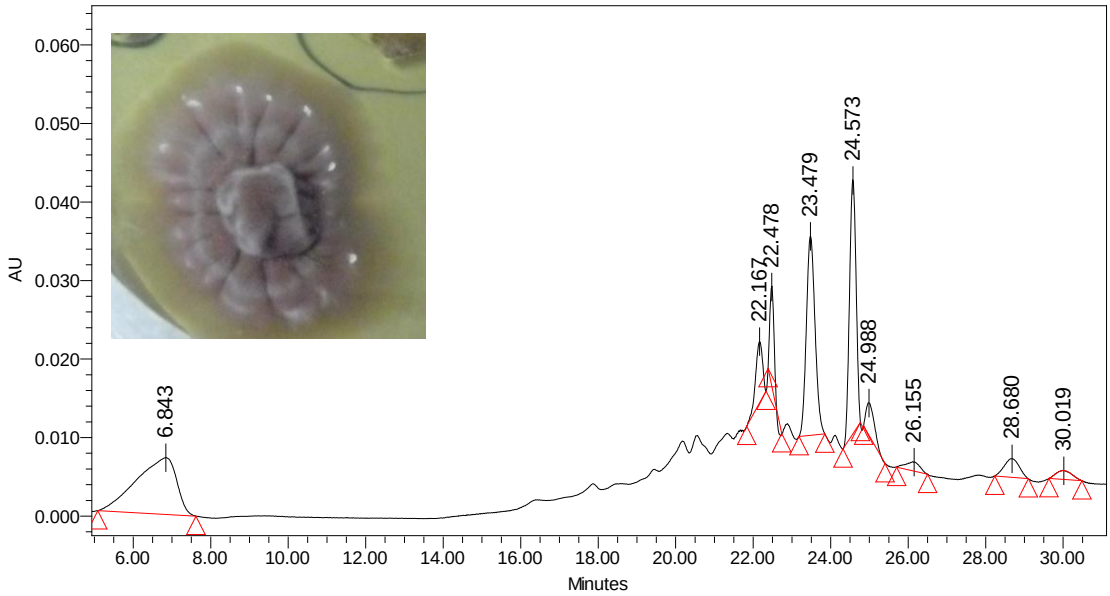


Figure 5.1 HPLC of *Cladonia uncialis* extract of a fungal culture in water and agar media. More than 9 secondary metabolites were detected when compared with the control media. The retention time of these traces are 6.80; 22.2; 22.5; 23.5; 24.6; 25.0; 26.1; 28.7 and 30.0. Image of *Cladonia uncialis* fungal grown in water and agar media is inserted.

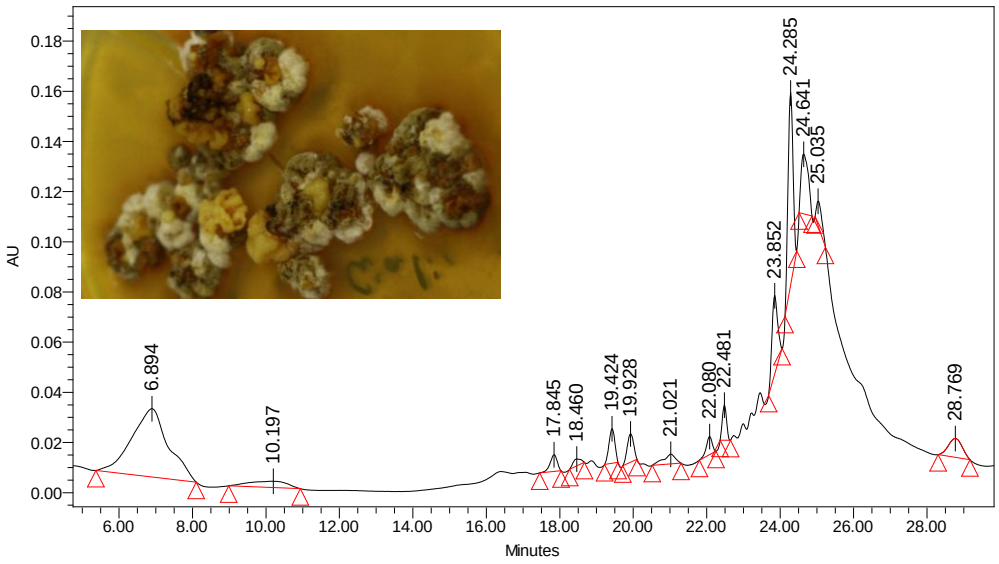


Figure 5.2 HPLC of *Cladonia uncialis* extract of a fungal culture in malt yeast sucrose media (10g/L sucrose) which shows 14 different secondary metabolites. Their retention times are 7.0; 22.8; 22.5; 23.8; 24.3; 24.6; 25.0; 23.8 28.8; 10.2; 17.8; 18.5; 19.4 and 19.9 minutes. Image of *Cladonia uncialis* fungal grown in malt yeast sucrose media is inserted.

The secondary metabolites profile changed when the fungal culture was grown on malt yeast extract medium with sucrose (Figure 5.2). The HPLC analysis of mycobiont that grown on yeast extract medium with sucrose revealed 5 new peaks (Figure 5.2). The retention time of these newpeaks are 10.2; 17.8; 18.5; 19.4 and 19.9 minutes.

The HPLC analysis of *Cladonia uncialis* fungal crude extract that was grown in malt-sucrose agar medium is shown in (Figure 5.3).

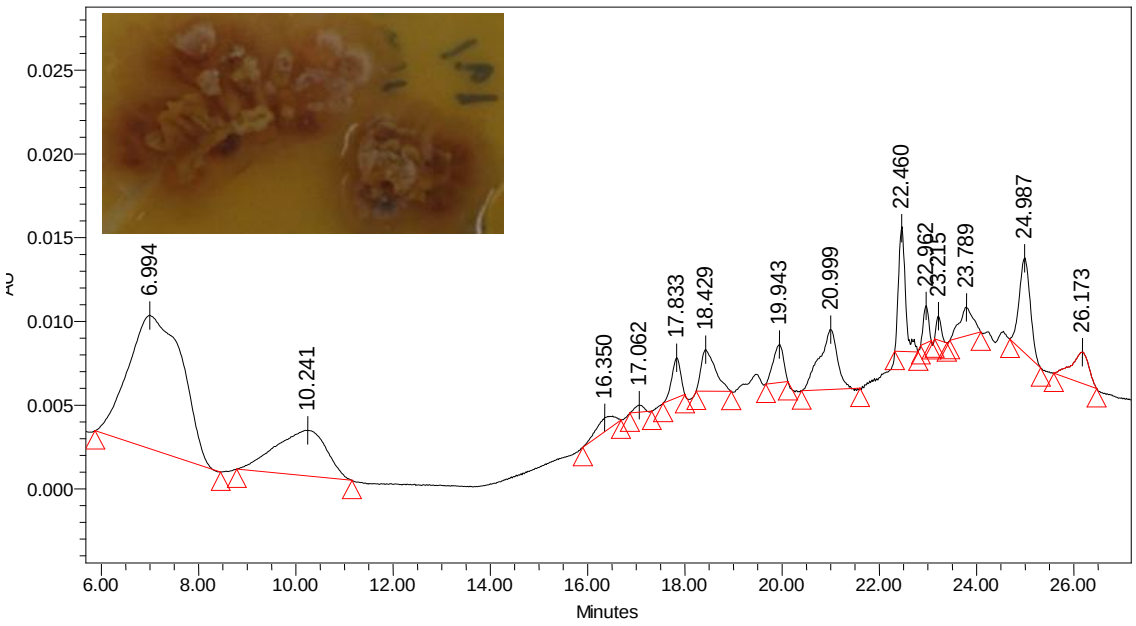


Figure 5.3 HPLC result of Cladonia *uncialis* extract of a fungal culture in malt agar sucrose media (10g/L sucrose). Image of *Cladonia uncialis* fungal grown in malt agar sucrose media is inserted.

The HPLC results from the fungal culture grown on trace elements medium with no sucrose added is shown in (Figure 5.4). The HPLC analysis revealed a major peak at 21.8 minutes with other minor peaks as in Figure 5.4. The NMR analysis for fungal culture extract grown on trace elements medium revealed signals consistent with a mixture of metabolites and could not be resolved.

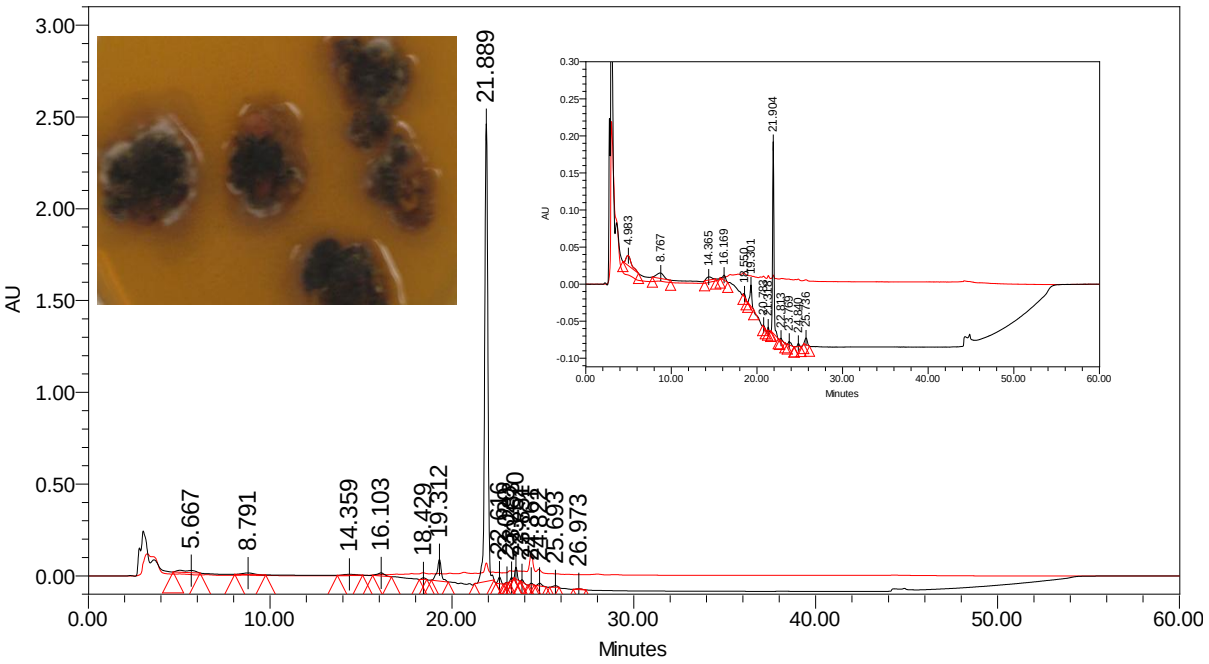


Figure 5.4 HPLC result of *Cladonia uncialis* extract of a fungal culture in trace element media which showed a large peak appeared at 21.9 minutes with other minor secondary metabolites. Other HPLC results which represent another replicate are also shown in the Figure. Image of *Cladonia uncialis* fungal grown in trace element media is inserted.

The HPLC analysis of the mycobiont crude extract that was grown on sucrose trace elements medium revealed a major peak at 22.0 minutes and a number of minor peaks (Figure 5.4). The NMR analysis for this crude extract could not be resolved due to the presence of signals which were consistent with a mixture of metabolites.

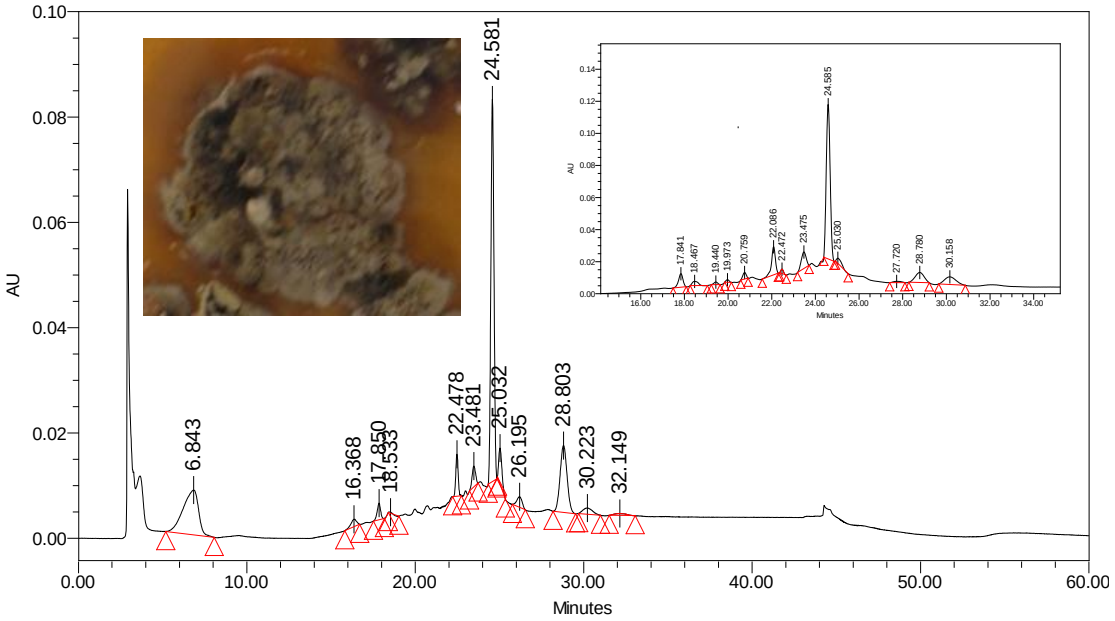


Figure 5.5 HPLC of a *Cladonia uncialis* extract of a fungal culture in trace element media with sucrose which show a large peak appear at 24.581 minutes with other minor secondary metabolites. The inset in the upper right of the figure shows another replica that gave the same result. Image of *Cladonia uncialis* fungal grown in trace element media with sucrose is inserted.

All the HPLC results from different fungal extracts were compared with HPLC results from different media extracts as control.

In this chapter the presence of different secondary metabolites in the culture which are not produced by the *Cladonia uncialis* natural lichen and the absence of usnic acid demonstrate that different culture conditions will affect the production of different secondary metabolites which is consistent with the 'one strain – many compounds' approach (Bode et al., 2002). The 'one strain – many compounds' approach suggests that the production of secondary metabolites occurs in response to a specific environmental or chemical stimulus within the habitat. By variation of culture conditions, for example via changing media composition, pH, aeration, temperature, radiation, et cetera, one may simulate these environmental stressors and thereby induce changes in the metabolic profile. Novel secondary metabolites synthesized by the organism could then be isolated and

its structure determined. This strategy may also be used in combination with genetic approaches for product isolation of new secondary metabolites.

5.6 Conclusion

In this chapter, the effect of different culture conditions on the production of different secondary metabolites including usnic acid in cultured lichen fungi *Cladonia uncialis* was investigated. Different media were used for this purpose and the crude extracts from all cultured fungi were analyzed by HPLC. The HPLC results showed that no usnic acid was produced in any of the cultured mycobionts while a wide range of secondary metabolites were detected in different crude extracts. Having sucrose in the media stimulated certain metabolites to be produced. Two different secondary metabolites (21.9 minutes in Figure 5.5 and 24.6 minutes in Figure 5.6) were also noticeable in the cultured mycobionts grown in trace elements medium and sucrose trace elements medium. However the chemical structure and the identification of all metabolites that appeared in the HPLC results was not successful for a number of reasons including the presence of more than one metabolite in the extracts and the low concentration of these metabolites.

The results in this chapter represent one strategy for changing the metabolic profile of the organism via changes in the culture condition which could be the way for having new products.

6 Chapter 6

The Identification of Secondary Metabolite Gene Clusters in the *Cladonia uncialis* Genome

6.1 Introduction

The advent of next-generation sequencing technology has revealed the biosynthetic possibilities of the genomic world. In searching for genes and their corresponding metabolites, researchers have discovered that organisms including bacteria, cyanobacteria, fungi, and plants, possess several more biosynthetic gene clusters than are accounted for in their metabolic profiles (Challis, 2008a; 2008b ; Gross, 2007). Genes with no corresponding metabolite have been referred to as 'orphan' genes (Gross, 2007). Many such genes have been identified throughout taxa and are putatively associated with multiple classes of molecules including terpenes, shikimic acid derivatives, polyketides, non-ribosomal peptides, lanthipeptides, and others (Pel et al., 2007; van den Berg et al., 2008; Udwary et al., 2007; Bentley et al., 2002; Ikeda et al., 2003; Oliynyk et al., 2007; Omura et al., 2001; Paulsen et al., 2005; Ziemert et al., 2014). The search for bioactive natural products for pharmaceutical and industrial use has been re-energized with the understanding that the

old process of extracting metabolites from cultures incubated under lab conditions has yielded only a tiny fraction of the organism's total biosynthetic potential (Koehn & Carter, 2005). 'Genome mining' for the identification and isolation of new natural products has consequently become a rapidly evolving and productive field of research (Nett et al., 2009; Bok et al., 2006; Corre & Challis, 2009; Challis, 2008a; 2008b; Gross, 2007; Wilkinson & Micklefield, 2007; Helfrich et al., 2014; Ziemert N et al., 2014).

In this chapter, I report that the biosynthetic potential of lichens far exceeds what may be observed through traditional surveys of natural products via extraction and isolation. A total of 56 putative biosynthetic gene clusters were identified through *de novo* whole-genome sequencing of the lichen, *Cladonia uncialis*. Gene clusters conform to major biosynthetic classes including non-ribosomal peptide synthases (NRPSs), polyketide synthases (PKSs), NRPS-PKS hybrid synthases, lanthipeptide synthases, and terpene synthases. *Cladonia uncialis* is remarkable for possessing diverse PKSs of multiple subclasses including iterative Type I highly reducing, partially reducing, and non-reducing synthases, as well as Type III PKSs. To our knowledge, this is the first comprehensive profiling of all secondary metabolite gene clusters within a species of lichen. Approaches for mining orphan genes and identifying new natural products are discussed.

6.2 Overview

6.2.1 Biosynthetic pathways involved in the synthesis of lichen secondary metabolites

Lichens are known for their ability to produce a variety of secondary metabolites via different biosynthetic pathways, (Figure 1.1). These pathways include polymalonate, shikimic acid and mevalonic acid pathways (Boustie and Grube 2005).

Most of the complex organic compounds isolated and identified from lichens are of fungal origin and produced via the polyketide pathway (Elix 1996). The key enzymes, polyketide synthases (PKSs), involved in the polyketide biosynthesis, are discussed in detail in Chapter 1, (1.2.3).

6.2.2 Non ribosomal peptides

Nonribosomal peptides are a class of peptides which have a wide range of biological activity, such as cytostatic, immunosuppressive, antibacterial, or antitumor properties. The discovery of the first non-ribosomal peptide was in 1968 when Gevers showed that the production of the cyclic decapeptide, gramicidin S, in cell extracts of *Bacillus brevis* occured even after the addition of RNase or ribosome inhibitors in the supernatant (Gevers *et al*., 1968). Additional biochemical analyses demonstrated that gramicidin S synthesis did not include tRNA molecules or aminoacyl-tRNA-synthetases. These compounds are produced by non-ribosomal peptide synthase (NRPS) enzymes in an assembly line mechanism. Non-ribosomal peptides are found in a variety of prokaryotic, cyanobacteria, low-

er order eukaryotes and sponges. Non-ribosomal modules often have the ability to incorporate nonproteinogenic amino acid into the synthesized natural product. Among the many advantages of the non-ribosomal peptides is the vast structural diversity of the compounds. This is due to the large numbers of the available substrates compared to the 20 amino acids available in ribosomal synthesis. There are over 300 different amino, hydroxyl, or carboxy-acid substrates that have been identified in the NRPS produced natural products (Schwarzer D *et al*., 2003). The unique amino acid and structures of non-ribosomal peptides offer useful and interesting properties such as antimicrobial, toxic and immunosuppressive activities. For this reason non-ribosomal peptides have been of particular interest to pharmaceutical and agrichemical companies and are the focus of many drug discovery initiatives (Schoenafinger and Marahiel 2009).

A typical NRPS module consists of an adenylation (A) domain, responsible for amino acid activation, a thiolation domain, also known as peptidyl carrier protein (PCP), which binds the activated amino acid and a condensation (C) domain that catalyzes peptide-bond formation. Additionally, a variety of optional domains have been described such as methyltransferase (MT) and epimerization (E) domains (Schwarzer D *et al*., 2003). The number of modules and their domain organization within NRPS enzymes controls the structures of the final product.

6.3 Material and Methods.

Taxonomic identification of *Cladonia uncialis*, culturing, isolation of the mycobiont, de novo genome sequencing of the mycobiont, and assembly of contigs from short reads were performed as described (Chapter 3, 3.2). Contigs were uploaded to the antiSMASH web server (Medema et al., 2011). Putative biosynthetic genes responsible for the formation of secondary metabolites, as well as accessory tailoring genes, were identified. Biosynthetic and tailoring genes (e.g. cytochrome P450s, terpene cyclases, prenyl transferases, methyl transferases, NAD(P)H-dependent oxidoreductases, and CoA/AMP ligases) were analyzed by BLAST searches against all known deposited genes in GenBank (Altschul et al., 1990). All the methods used in this chapter are summarized in Figure 6.1.

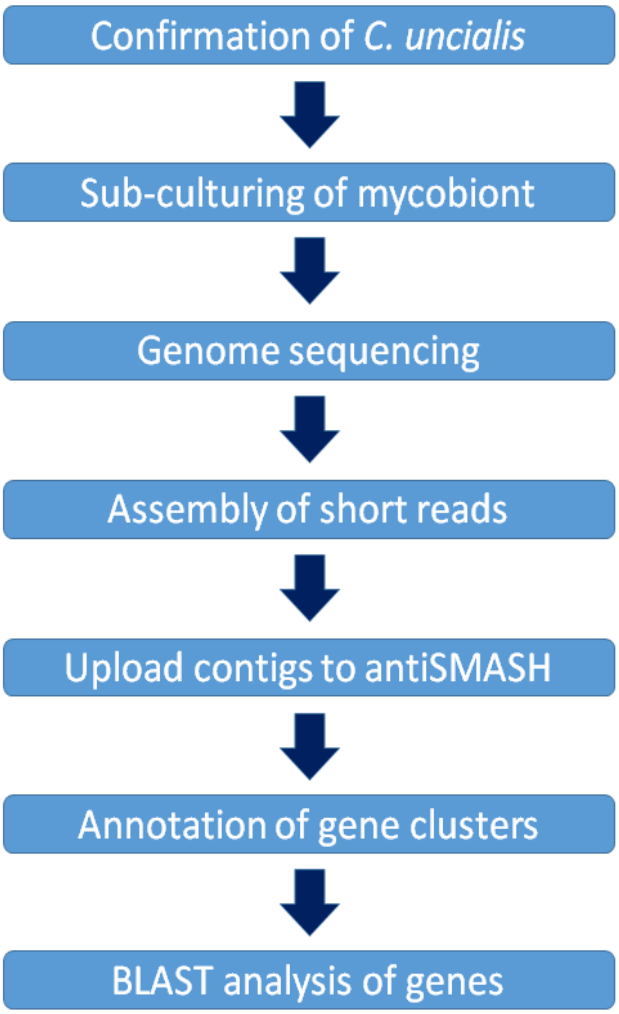


Figure 6.1 Summary of gene annotation methods.

6.4 Results and discussion

6.4.1 Features of secondary metabolite biosynthetic gene clusters

Raw DNA sequences obtained from whole-genome de novo sequencing of the mycobiont of *Cladonia uncialis* were assembled into 2271 contigs of which 1859 were at least 2Kb in length. Fifty six putative secondary metabolite biosynthetic gene clusters were identified by using antiSMASH to scan the genome and identify sequences that had homology to known genes. Identified gene clusters were predicted to be involved in the biosynthesis of polyketides, non-ribosomal peptides, terpenes, lanthipeptides, and polyketide-nonribosomal peptide hybrids. A striking feature of the *Cladonia uncialis* genome is the diversity of polyketide synthases (PKSs), which included highly, partial, and non-reducing iterative type I PKSs, type III PKSs, and hybrid type I PKS-NRPSs.

Many lichen metabolites are produced exclusively by the fungal partner in the lichen association. To date, no biosynthetic gene has been linked with specific metabolites by functional expression in the lichen fungus. For these reasons, bioinformatics via antiSMASH and BLAST failed to identify putative metabolites or specific catalytic functions from any of the 56 genes identified. This preliminary work is therefore intended to identify, annotate, and showcase for the first time the total genetic capacity of lichen fungi for the biosynthesis of secondary metabolites. The state of expression or function of synthetic genes and identified tailoring enzymes are unknown. In some cases, genes in

proximity to an identified synthetic gene will have no apparent tailoring function or are possibly unrelated to the biosynthetic process. Successful genome-based approaches employed in the discovery of novel natural products in organisms other than lichens will be described. The utility of this research as well as future genome annotation efforts leading to the identification of novel natural products will therefore rest in the task of using these and other techniques to isolate products.

6.4.2 NRPS gene clusters in the *Cladonia uncialis* genome

A total of ten NRPS gene clusters were identified, consecutively named CU-NRPS-1 to CU-NRPS-10. Two genes, CU-NRPS-1 and CU-NRPS-2, were determined to be linear NPRSs possessing four or five distinct catalytic modules and with extensive oxidative and reductive tailoring genes in proximity, suggesting that both genes are responsible for the biosynthesis of complex peptide products (Figure 6.3). ATP-binding cassette (ABC) transporters, ATP-dependent transmembrane proteins that transport compounds (Jones & George, 2004), were also found within the gene clusters of both these genes. These may be coincidently proximal to these synthetic genes or may signify that a physiological role of the metabolite produced by CU-NRPS-1 and CU-NRPS-2 involves translocation across the cell membrane.

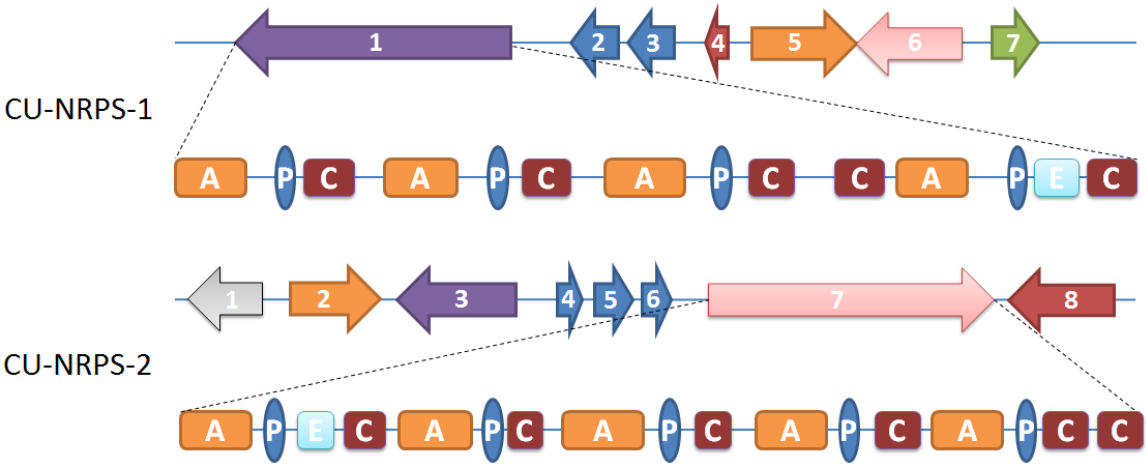


Figure 6.2 Organization and contents of the biosynthetic gene clusters of CU-NRPS-1 and CU-NRPS-2. (Top): (1) CU-NRPS-1; (2,3) ATP-binding cassette (ABC) transporter; (4,7) cytochrome p450 oxidative enzyme (5) 3-oxoacyl synthase containing ketoacyl reductase, beta-ketoacyl synthase, unspecified n-terminal domain; (6) unspecified putative fatty acid synthase beta subunit dehydratase. (Bottom): (1) cytochrome p450 oxidative enzyme (2) malonyl CoA-acyl carrier protein transacylase; (3,4) Short-chain dehydrogenase / reductase (SDR); (5,6) ABC transporter; (7) CU-NRPS-2; heterokaryon incompatibility protein. (Abbreviations): (A) adenylation domain; (P) peptidyl carrier domain; (C) condensation domain; (E) Epimerization domain.

Eight short non-canonical NRPS genes were also identified (Table 6.1). The domain architecture of these genes are identical, notably lacking a condensation domain, but possessing a single adenylation domain, a peptidyl carrier domain, and a thioesterase (Figure 6.3). The composition of these genes suggest the gene products synthesize simple, possibly single reaction, peptides. Two reductive genes were associated with (CU-NRPS-3), suggesting reduction of a carbonyl-peptide, and an O-methyltransferase was found proximal to one of the genes, suggesting methylation of a hydroxyl-peptide. A gene with sequence homology to a gene of unknown function sequenced in *Penicillium chrysogenum* (Accession no. XM_002556773) was found adjacent to (CU-NRPS-4).

NAME	ORGANIZATION	ACCESSORY GENES
CU-NRPS-3	A-P-TE	1- Enoyl reductase 2- dehydrogenase/reductase
CU-NRPS-4	A-P-TE	1- Upstream gene which consists of 1752 amino acid sequence. This gene is similar to Penicillium chrysogenum Wisconsin 54-1255 fungi 2- O-methyltransferase gene
CU-NRPS-5	A-P-TE	No gene cluster
CU-NRPS-6	A-P-TE	No gene cluster
CU-NRPS-7	A-P-TE	No gene cluster
CU-NRPS-8	A-P-TE	No gene cluster
CU-NRPS-9	A-P-TE	No gene cluster
CU-NRPS-10	A-P-TE	No gene cluster

Table 6.1 Simple non-canonical non-ribosomal peptide synthetases identified in *Cladonia uncialis.*

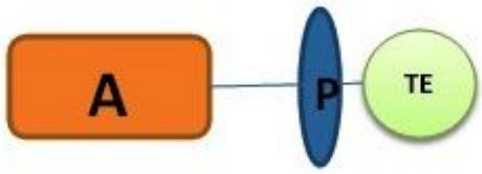


Figure 6.3 Domains architecture of noncanonical NRPS gene.

6.4.3 PKS–NRPS Hybrid gene clusters

PK-NRP hybrids combine acetate-derived polyketide chains with the more than 300 proteinogenic and nonproteinogenic amino acids that could be incorporated by NRPS modules (Schwarzer D *et al*., 2003). The PKS starts building the molecule in an iterative way and then passes it to the NRPs to add to it one or more amino acids (proteinogenic and

nonproteinogenic). The ketosynthase domain (KS) attaches a malonyl-CoA extender unit to an acetyl-CoA starter molecule by carbon-carbon bond formation via Claisen condensation reaction (Christopher, R. et al., 2010, Strieker, M., 2010). The acyl transferase (AT) domain supports loading onto the acyl carrier protein (ACP) and the KS domain, respectively. Optionally, each elongation step can be followed by further ß-keto processing steps, mediated by ketoreductase (KR), dehydratase (DH), and transacting enoyl reductase (ER) domains. Furthermore, optional methyltransferase domains (CMet) may introduce a methyl group into the polyketide chain. After a certain number of elongation steps, the polyketide intermediate is then transferred to the NRPS module. The amino acid is selected and activated by the adenylation domain (A) and then transferred onto the thiolation domain (T). The condensation domain (C) finally links the polyketide chain to the activated amino acid (Boettger, D. *et al*., 2012, Christopher, R. *et al.*, 2010, Eisfeld, K., 2009). The hybrid chain is then off-loaded by means of a terminal domain.

Four PKS-NRPS hybrid gene clusters were identified, three of them without tailoring enzymes (Figure 6.4). The first hybrid, third and fourth hybrids, CU-PKS-NRPS-1, CU-PKS-NRPS-3 and CU-PKS-NRPS-4 presents with a partially reducing PKS component possessing an enoyl reductase that presumably operates on an alkene intermediate possibly came from the NRPS component possessing a ketoreductase (KR). It is unusual that ketoreductase and dehydratase domains, necessary for generating alkene intermediates

for processing by the ER domain, are not present within the PKS, though it is further unusual for KR domains to be present within the NRPS. This may suggest that both genes cooperate to reduce the nascent product of the NRPS components. All four NRPS genes in the hybrid PKS-NRPS genes possess a complete C-A-P catalytic module necessary for at least one round of catalytic operation and three of them possess an unspecified terminal domain presumably involved in release of the product from the enzyme (Figure 6.4). The absence of an ACP domain in the PKS components of all four PKS-NRPS hybrids suggests that polyketide synthesis may proceed without the ACP domain or instead relies on a domain from the NRPS such as the peptidyl carrier domain (P domain) which functions analogously to an ACP during peptide elongation or using the phosphopantetheinyl (PP or ACP) arm directly from acetyl CoA as in type III PKS genes. One sequence of the accessory genes associated with CU-PKS-NRPS-1 is similar to the sequence of an unspecified drug resistance transporter suggesting that this metabolite is involved in defence against toxins. The gene architecture of CU-PKS-NRPS-2 suggests that a fully reduced polyketide is incorporated into the final product.

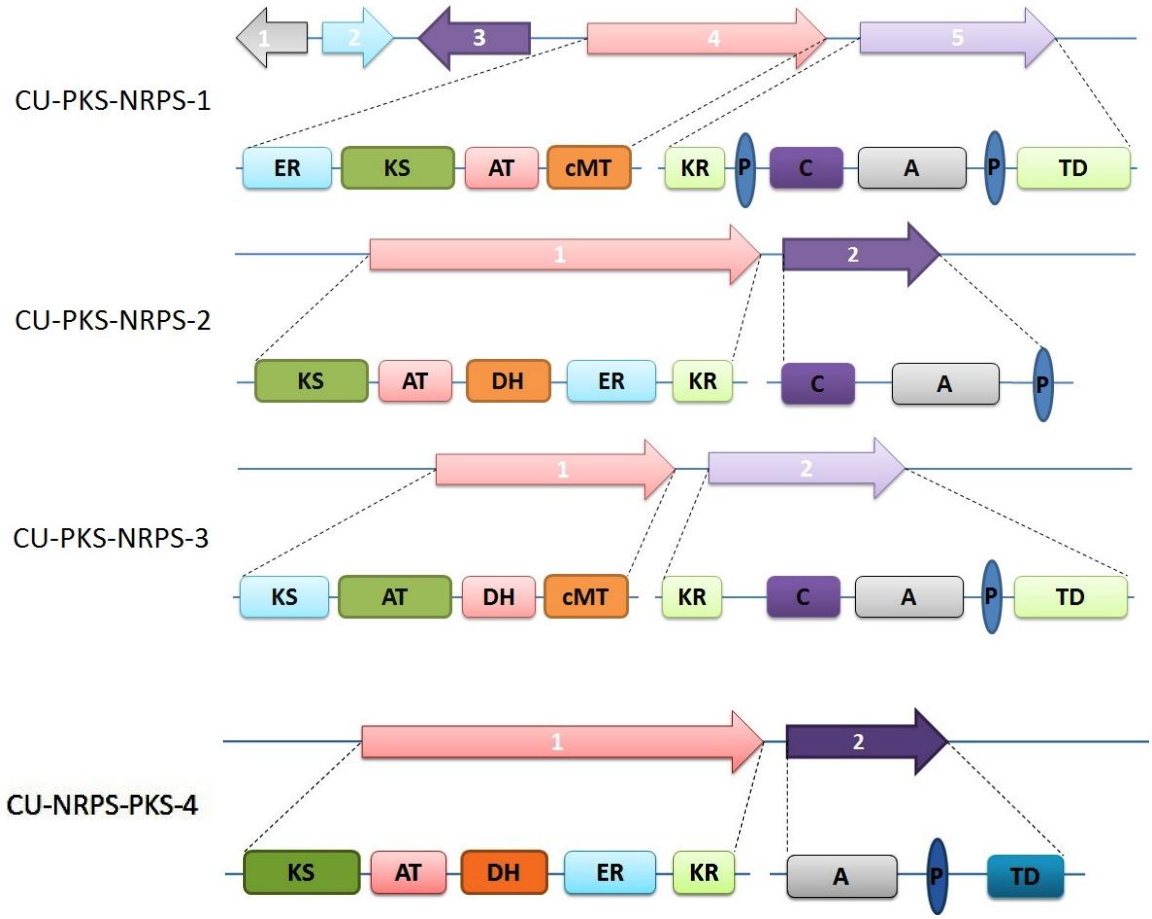


Figure 6.4 Contents and domain architecture of four PKS-NRPS hybrid synthases. (CU-NRPS-PKS1) (1) enoyl reductase; (2) drug resistance transporter, unspecified; (3) FAD-linked oxidative protein; (4) PKS component; (5) NRPS component. (CU-NRPS-PKS2) (1) PKS component; (2) NPRS component. (CU-NRPS-PKS3) (1) PKS component; (2) NRPS component (Abbreviations) ER, enoyl reductase; KS, ketosynthase; AT, acyltransferase; CMet, C-methylation domain; KR, ketoreductase; P, peptidyl carrier protein; A, adenylation domain; C, condensation domain; TD, terminal domain, unspecified; DH, dehydratase. (CU-NRPS-PKS4) (1) PKS component; (2) NRPS component.

It is known that PK/NRP hybrid natural products have a large number of biological activities such as epothilone which has anticancer activity, Rapamycin which is immunomodulatory and Pristinamycin IIA which is antimicrobial. Functional expression of these

genes might be used for industrial and medical fields (Boettger, D. *et al*., 2012, Christopher, R. *et al.*, 2010).

6.4.4 PKS gene clusters

Whole-genome sequencing of the mycobiont of *C. uncialis* identified a total of 32 iterative type I PKS genes in addition to two type III PKS gene clusters. The iterative type I PKS genes were sub-categorized into 16 non-reducing, one partially reducing, and 15 fully-reducing PKSs. Approximately half of all PKSs identified were associated with tailoring enzymes, most commonly oxidative enzymes with sequences being cytochrome p450 oxidative enzymes or flavin adenine dinucleotide (FAD)-dependent oxidases. The number of the PKS genes identified in the *Cladonia uncialis* genome is much larger when compared with other lichen genomes. For example, the number of PKS genes that were detected in the lichen-forming fungus *Endocarpon pusillum* was 14 PKS in addition to 2 NRPS genes (Wang Y. *et al*., 2014). These enzymes produce polyketides through iterative or linear decarboxylative Claisen condensation cycles between malonyl-CoA extension units and a growing polyketide chain (Keatinge-Clay, 2012; Tsai and Ames, 2009; Smith and Tsai, 2007, Khosla *et al*., 2007, Staunton and Weissman, 2001). Other extension units, e.g., methylmalonyl-CoA, may also be accepted by some PKSs (Chan *et al*., 2009). Out of the 32 PKS genes, 16 highly reducing and partially reducing PKS genes were identified (Figure 3.6). To date only two highly reducing PKS genes were identified in lichen fungi (Kim J. A., *et al*., 2012, Wang Y. *et al*., 2011). The functions of all the 32 PKS genes were not identified by the BLAST search since lichen PKS genes have not

been linked to any metabolites. The presence of saturated carbon side chains in some of the lichen metabolites (Figure 6.5) suggested that these highly reducing PKS genes might be involved in their production.

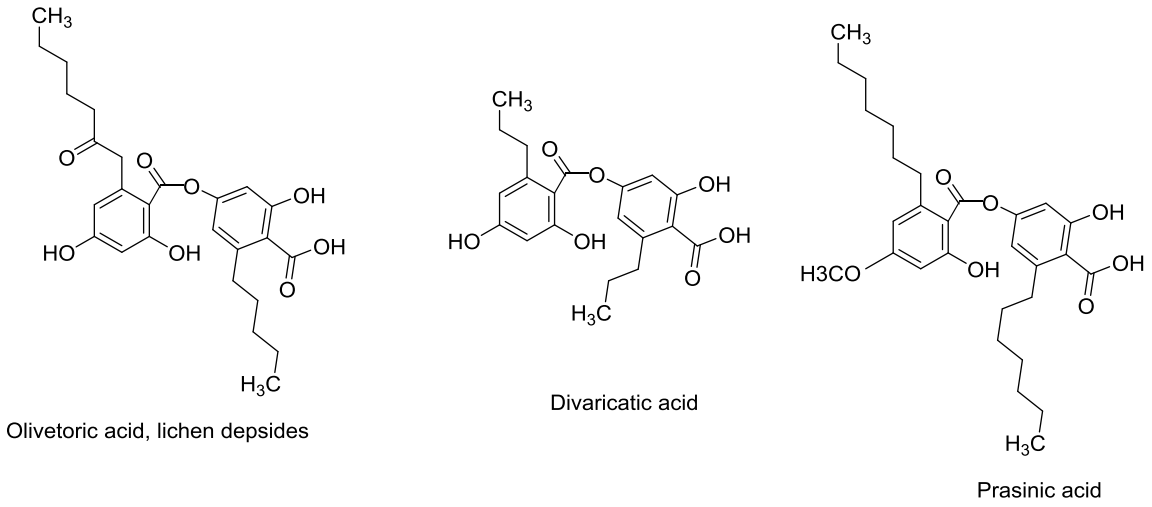


Figure 6.5 Lichen metabolites which have saturated carbon chains as PKS starter units.

The N terminal domain in PKS genes (SAT domain) can accept different starter units that are more complex than acetyl CoA, especially when they are located at the same cluster of other PKSs or associated with other biosynthetic systems, such as non-ribosomal peptide synthetases (NRPSs), FASs, and terpene biosynthesis enzymes (Crawford and Townsend, 2010). Our results agree with this theory since there are 3 highly reducing PKS genes in *Cladonia uncialis* that cluster with type III PKS genes, non-reducing PKS genes and terpene synthase genes in their clusters (Figure 6.6). A large variety of genes of tailoring enzymes such halogenation, O-methylation, C-methylation, oxidation, reduction and decarboxylation were found associated with 11 highly reducing PKS genes (Figures 6.6, 6.7, 6.8 and 6.9). The presence of this variety of enzymes in the gene clusters indi-

cates that the resulting polyketides likely possess remarkable structural diversity. For example, the product of CU-PKS-11 is suspected to undergo post-elongation modifications including halogenation, O-methylation, mono-oxygenation, and reduction. It has been observed that an accessory gene of CU-HRPKS 2 suggests that this polyketide undergoes N-acylation though no other tailoring enzymes were identified suggesting addition of an amino or other nitrogen-containing group. An aminotransferase was instead found within the gene cluster of CU-HRPKS-4 as well as multiple oxidative and reductive elements. Interestingly, within this cluster a putative crotonyl-CoA reductase was found. This enzyme has been implicated in the generation of methylmalonyl-CoA (Li *et al.*, 2004; Liu and Reynolds, 2001), suggesting that this PKS may biosynthesize polyketides via condensation of methylmalonyl-CoA extension units. In addition to another putative crotonyl-CoA reductase gene within the gene cluster of CU-HRPKS-4, a carboxymuconolactone decarboxylase was also discovered. Although the accessory enzymes present do not appear to support a biosynthetic pathway to carboxymuconolactone from a polyketide but the presence of this gene suggests that this product may be structurally similar to carboxymuconolactone or at the very least undergoes decarboxylation. Within the CU-HRPKS-6 gene cluster a gene homologous to 8-amino-7 oxononanoate synthase was discovered. This enzyme binds L-alanine and 6-carboxyhexanoyl-CoA, to form 8-amino-7 oxononanoate, a committed precursor of biotin (vitamin B_7) biosynthesis (Webster et al., 2000). 6-carboxyhexanoyl-CoA could be produced through three condensation cycles using a fully-reducing PKS consistent with the gene architecture of CU-HRPKS-6.

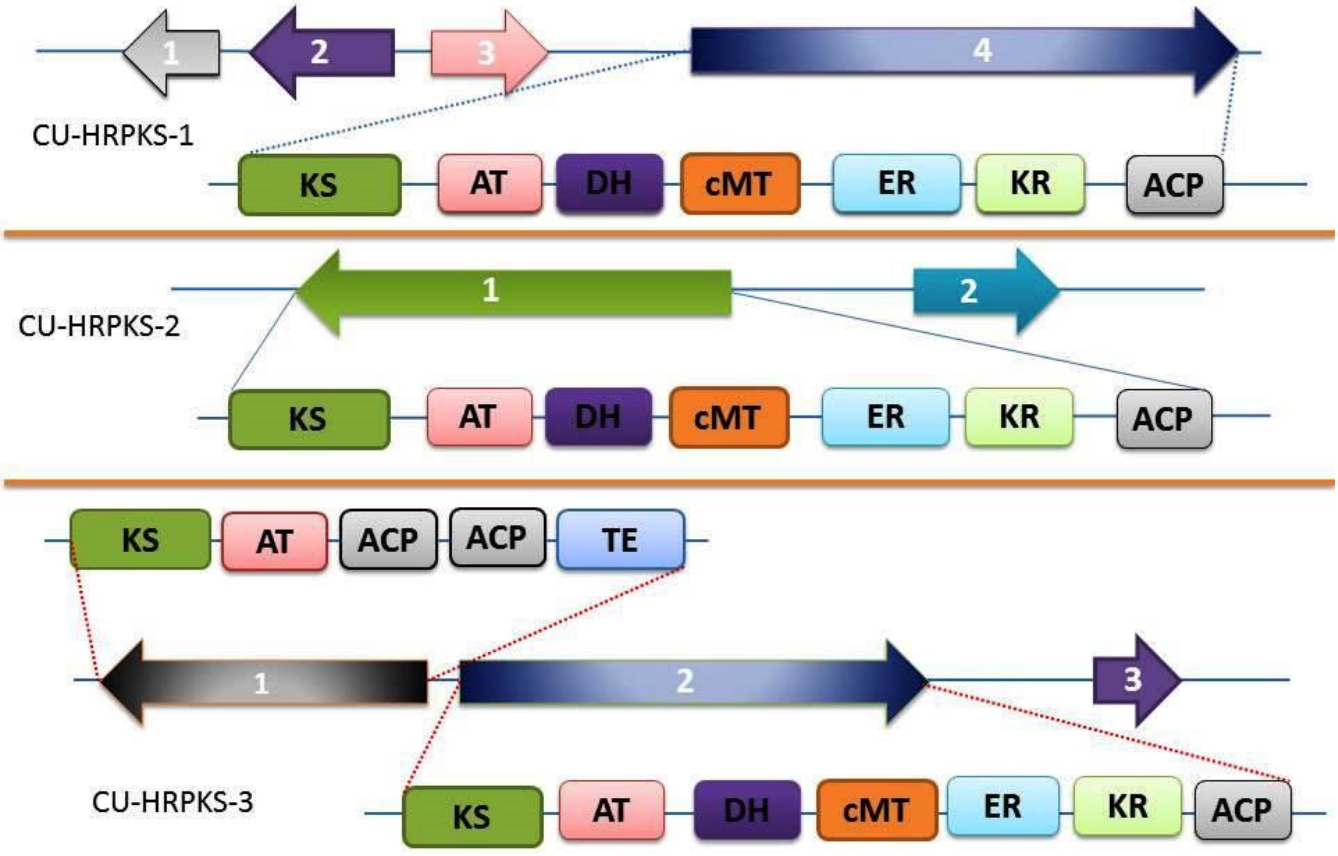


Figure 6.6 Contents and domain architecture of three HRPKS genes. (CU-HRPKS-1) (1) Short-chain dehydrogenase/reductase; (2) cytochrome P450; (3) Terpene synthase; (4) Highly reducing PKS gene. (CU-HRPKS-2) (1) Highly reducing PKS gene; (2) Type III PKS gene. (CU-HRPKS-3) (1) NRPKS gene; (2) HRPKS gene; (3) Halogenase. These genes are also shown in Figure 3.6.

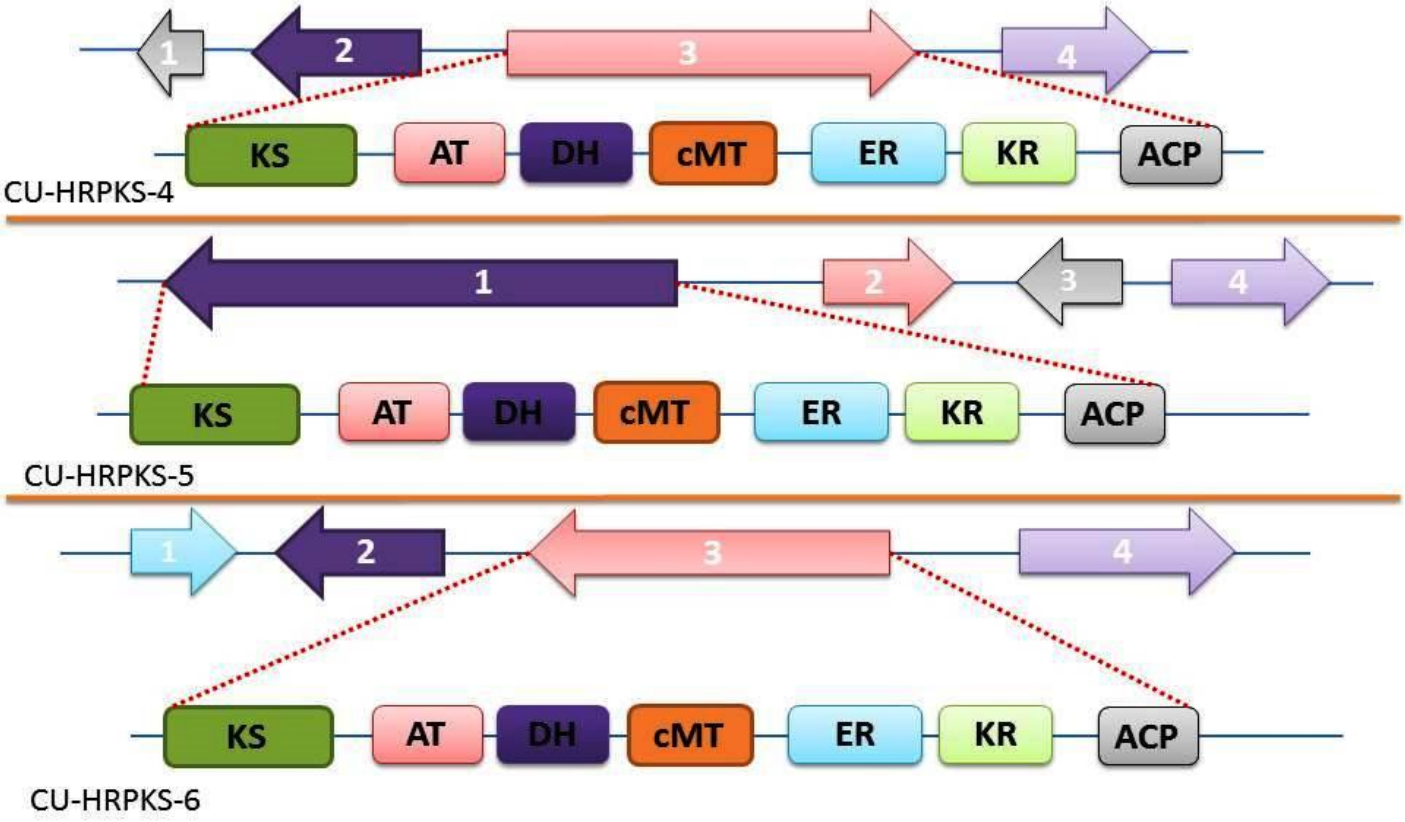


Figure 6.7 Contents and domain architecture of three HRPKS genes. (CU-HRPKS-4) (1) putative carboxymuconolactone decarboxylase; (2) crotonyl-CoA reductase / alcohol dehydrogenase; (3) HRPKS; (4) crotonyl-CoA reductase / alcohol dehydrogenase. (CU-HRPKS-5) (1) HRPKS; (2) FAD linked oxidase domain protein; (3) cytochrome P450; (4) FAD linked oxidase domain protein. (CU-HRPKS-6) (1) Drug resistance transporter; (2) 8-amino-7-oxononanoate synthase; (3) HR PKS gene; (4) sugar transport protein. These genes are also shown in Figure 3.6.

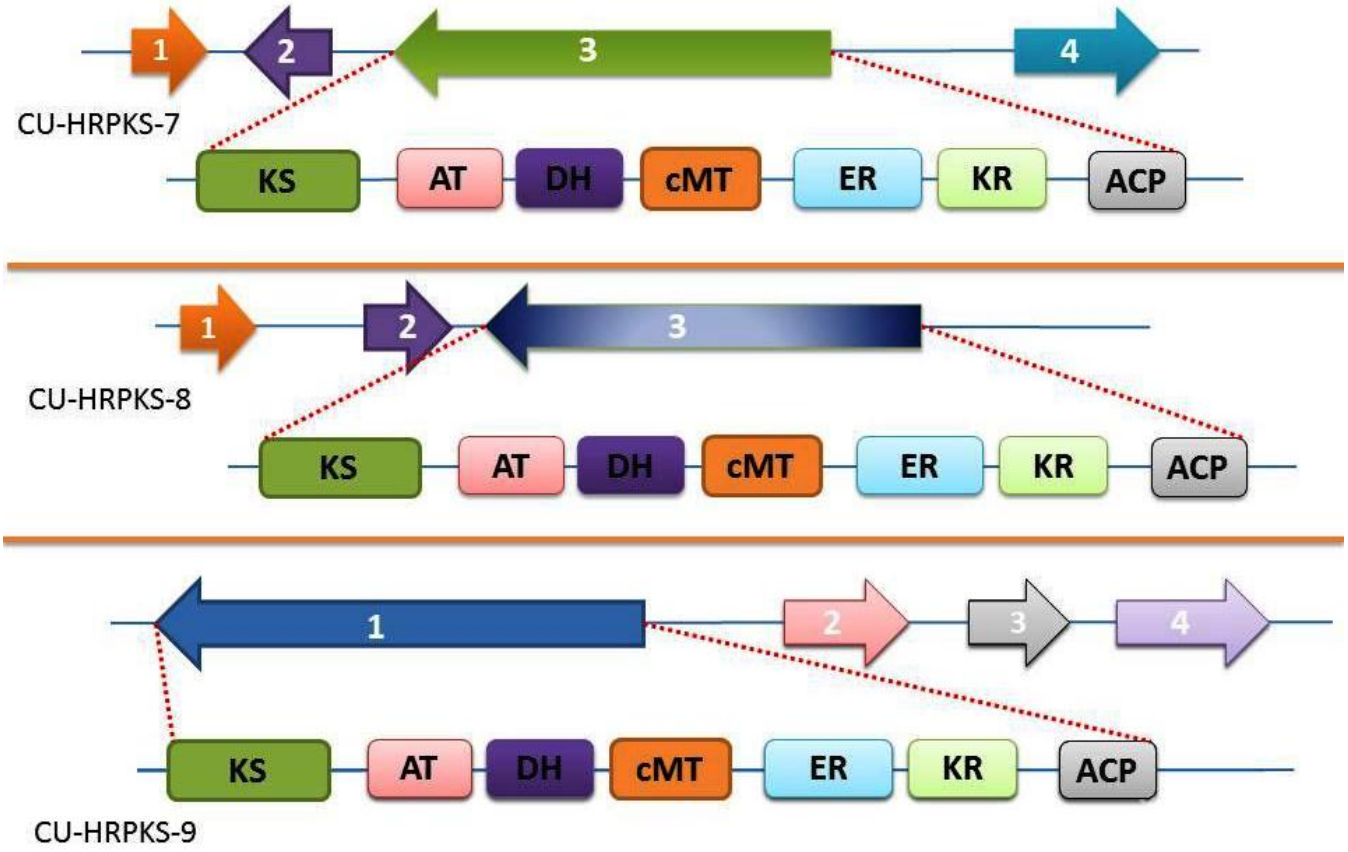


Figure 6.8 Contents and domain architecture of three HRPKS genes. (CU-HRPKS-7) (1) short-chain dehydrogenase/reductase; (2) Cytochrome P450 (3)HRPKS gene; (4) NAD-dependent epimerase/dehydratase. (CU-HRPKS-8) (1) Major facilitator superfamily MFS ((2) O-methyltransferase; (3) Highly reducing PKS gene. (CU-HRPKS-9) (1) HRPKS gene; (2) O-methyltransferase; (3) Short-chain hydrogenase/reductase ; (4) FAD linked oxidase domain protein. These genes are also shown in Figure 3.6.

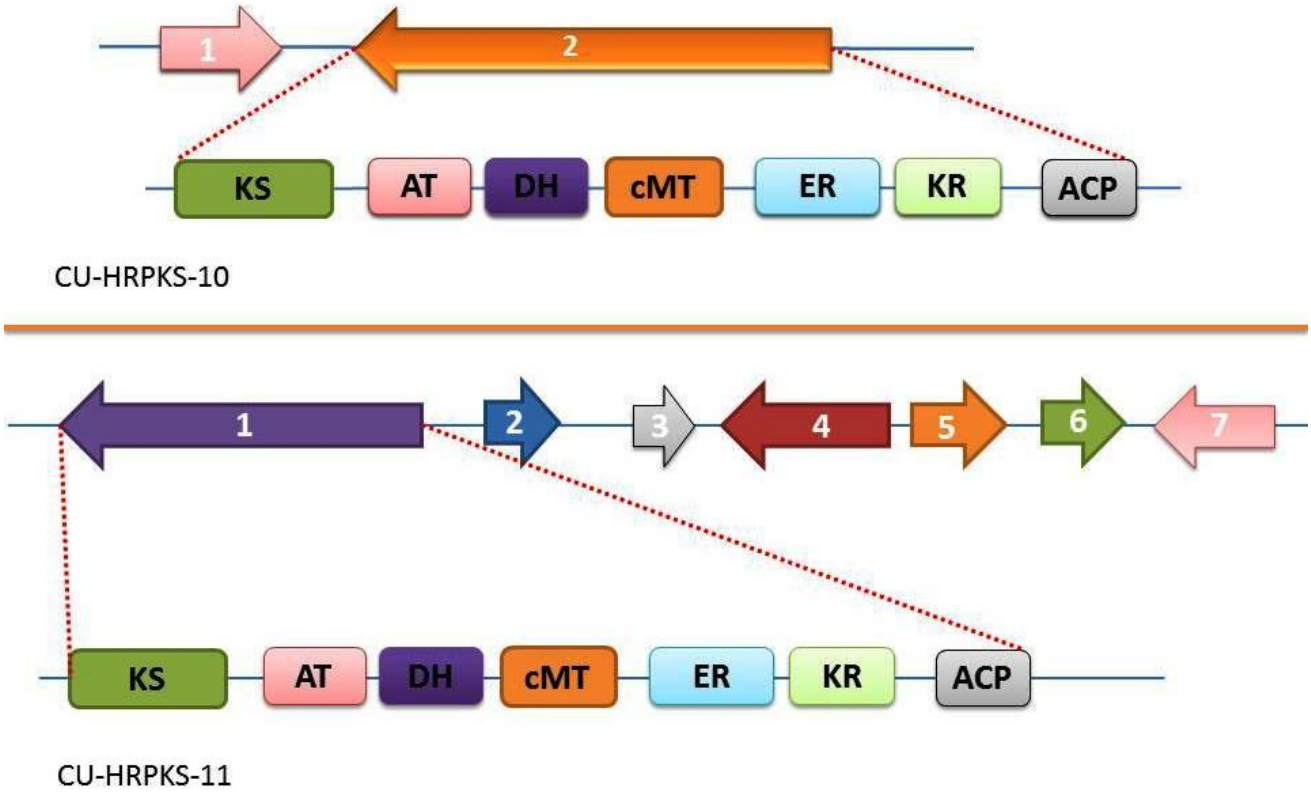


Figure 6.9 Contents and domain architecture of two HRPKS genes. (CU-HRPKS-10) (1) Methyltransferase; (2) HRPKS gene. (CU-HRPKS-11) (1) HRPKS gene; (2) cytochrome P450; (3) halogenase (4) cytochrome P450; (5)short-chain dehydrogenase/reductase SDR; (6) O-methyltransferase; (7) short-chain dehydrogenase/reductase SDR. These genes are also shown in Figure 3.6.

Sixteen non reducing PKS genes were identified in the *Cladonia uncialis* genome sequence. The accessory genes of 8 NRPKS genes out of 16 were identified by using antiSMASH followed by manual annotation. Out of these 8 genes only one is thought to be the usnic acid PKS gene (Chapter 3); another gene is thought to produce halogenated anthraquinone polyketide (Chapter 4) and one non-reducing PKS gene is located head to head with a highly reducing PKS gene (Figure 6.6). Annotation of 5 NRPKS genes with their accessory genes are presented in Figures 6.10 and 6.11.

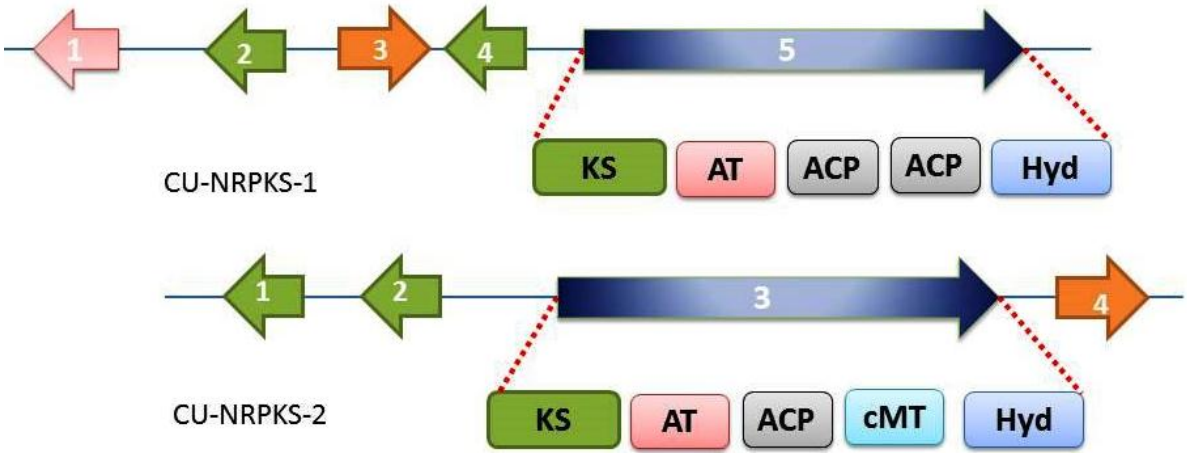


Figure 6.10 Contents and domain architecture of two NRPKS genes. (CU-NRPKS-1) (1) O-methyltransferase; (2) FAD linked oxidase domain protein; (3) aldo/keto reductase family oxidoreductase; (4) FAD linked oxidase domain protein; (5) NRPKS gene. (CU-NRPKS-2) (1) cytochrome p450; (2) cytochrome p450, (3) NRPKS gene; (4) N-Acyltransferase.

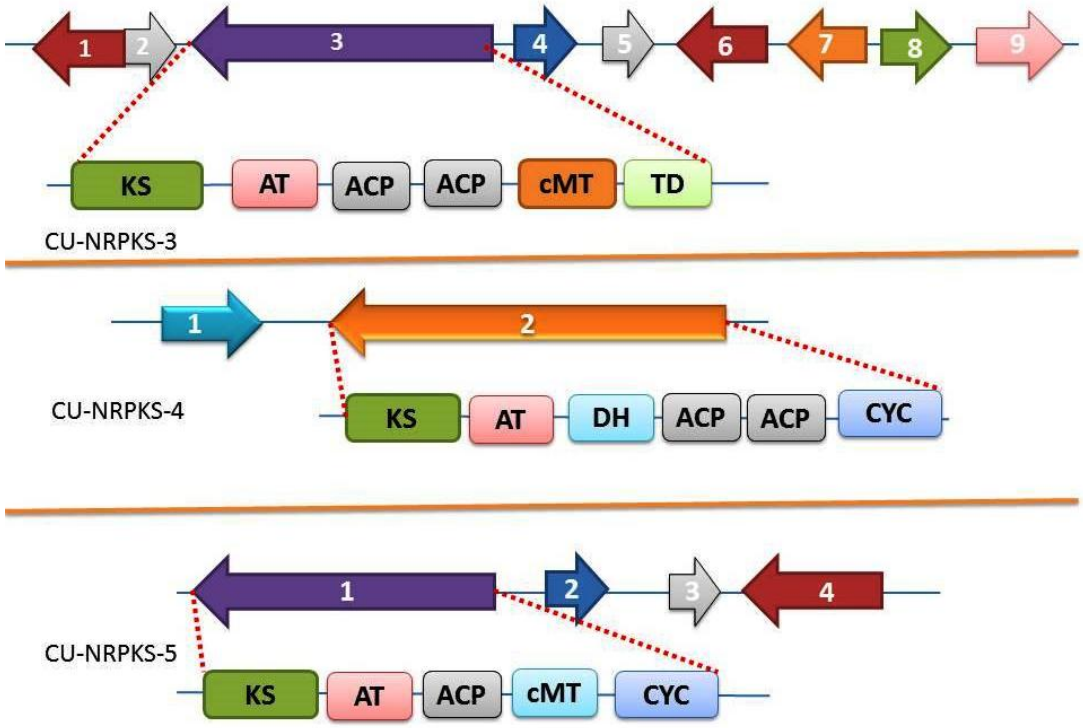
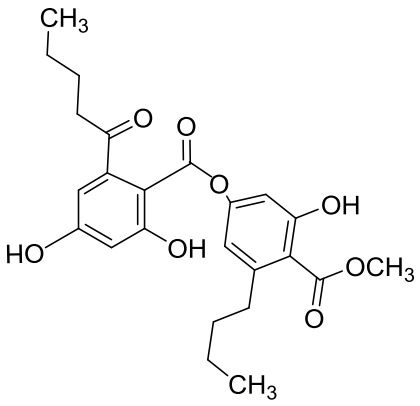


Figure 6.11 Contents and domain architecture of three NRPKS genes. (CU-NRPKS-3) (1) ABC transporter ATP-binding protein; (2) aminotransferase; (3) NRPKS gene; (4) crotonyl-CoA reductase / alcohol dehydrogenase; (5) FAD linked oxidase domain protein; (6) Drug resistance transporter, EmrB/QacA; (7) FAD linked oxidase domain protein; (8) major facilitator transporter; (9) Short-chain dehydrogenase/reductase SDR. (CU-NRPKS-4) (1) O-methyltransferase; (2) NRPKS gene. (CU-NRPKS-5) (1) Cytochrome P450; (2) NADPH- dependent oxireductase; (3) sugar transport protein.

Type III PKSs represent simple biosynthetic enzymes that found in plants, bacteria, and fungi (Austin and Noel, 2003; Funa *et al*., 2006; Ferrer *et al*., 1999). They are represented by the gene families of chalcone synthases (CHSs) and stilbene synthases (STSs). They consist of small ketosynthases that use CoA esters as substrates and synthesize polyketides by using a single catalytic center. A type III PKS gene is composed of a single polypeptide chain lacking the phosphopantetheinyl (PP or ACP) arm, which is typically present in type I and type II PKSs. Both CHS and STS select either p-coumaroyl-CoA or cinnamoyl-CoA as a starter unit and carry out successive condensations with three molecules of malonyl-CoA. The presence of a type III PKS in *Neurospora crassa*, that catalyzes the synthesis of 2'-oxoalkylresorcylic acids, suggests that a similar type III PKS could be involved in the biosynthesis of some coupled phenolics with alkylresorcinol moieties (Funa *et al*., 2007). Coupled phenolics with oxoalkylresorcylic acid moieties are found occasionally, in lichens. As an example, the depsidone loxodin is a lichen metabolite which has oxoalkylresorcylic acid moieties and could be produced by a type III PKS gene in cooperation with a type one highly reducing PKS gene (Figure 6.12).

In this Chapter, two Type III PKS gene clusters are detected in the *Cladonia uncialis* genome which might explain the presence of coupled phenolics with oxoalkylresorcylic ac-

id moieties in lichen fungi. One of these genes has a HRPKS gene in its cluster (Figure 6.6) while no accessory genes were identified for the other gene.



loxodin

Figure 6.12 Examples of lichen depsidones with oxoalkylresorcylic acid moieties. Depsidone (loxodin) with a 1'-oxoresorcylic acid moiety.

Lanthipeptide synthase

Lanthipeptides are a class of ribosomally synthesised peptide natural products. The lanthionine structure consists of two alanine residues that are linked through a thioether that connects their ß-carbons. Lanthipeptides are characterized by lanthionine residues in which thioether bridges link cysteine and serine/threonine residues (Yu *et al*., 2013). In a first step of lanthipeptide biosynthesis both of Ser and Thr residues are dehydrated to form dehydroamino acids. Then the double bond in these dehydroamino acids is subsequently coupled to the thiol group of Cys residues. Lanthipeptides have a range of different activities such as antimicrobial, morphogenetic and antiviral (Dischinger, J. et al., 2014, Willey JM *et al*., 2007, Schnell *et al*., 1988). A single lanthipeptide synthase gene without adjacent accessory genes was identified in the genome of *Cladonia uncialis.*

Intriguingly, and to the best of our knowledge, no lanthipeptide secondary metabolites have been isolatedfrom lichen fungi to date. BLAST analysis of the synthetic gene failed to identify putative metabolites or other identifying features. For this reason a gene product could not be speculated.

6.4.5 Terpene synthase

Terpenes are the largest group of natural products. Approximately 25,000 reported terpene structures are assembled from the basic unit isoprene (Gershenzon and Dudareva, 2007). Members of this family include the monoterpenes, sesquiterpenes, diterpenes, steroids, carotenes and polyprenoids such as the ubiquinone side chain. Terpenoids are extraordinarily diverse but they all originate through the condensation of a phosphorylated derivative of hemiterpene, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) giving geranyl pyrophosphate (GPP) (Figure 6.13) (Spurgeon and Porter, 1981). In higher plants, IPP is derived from the classic mevalonic acid pathway (Bohlmann *et al*., 1998). A classification of the terpenes has been established based on the number of isoprene (or isopentane) units incorporated in the basic molecular skeleton. For examples monoterpenes contain 2 isoprene units, sesquiterpenes contain 3 isoprene units, diterpenes contain 4 isoprene units, sesterterpenes contain 5 isoprene units, triterpenes contain 6 isoprene units, carotenoids contain 8 isoprene units and rubber contains more than 100 isoprene units.

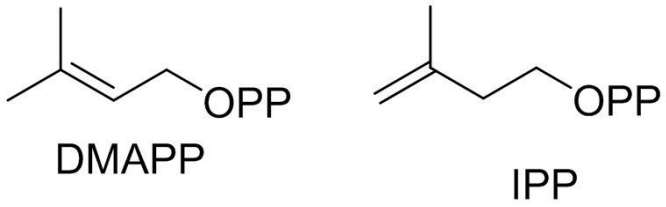


Figure 6.13 The bulding blocks of terpenes isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP).

Terpenenes have been identified previously in lichens. Examples of lichen terpenes are diterpene (16α-hydroxykaurane), sesterterpene (Retigeranic acid), riterterpene (Zeorin), carotenoids (Zeaxanthin) (Figure 6.14)

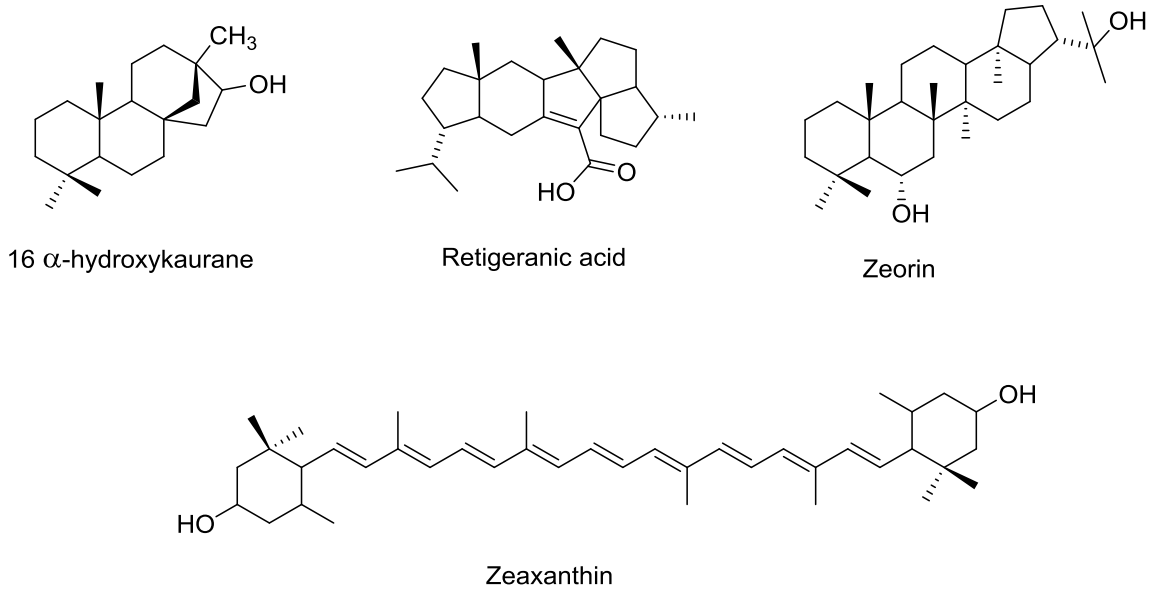


Figure 6.14 Examples of lichen terpenenes.

In the present work a total of seven terprene biosynthetic gene clusters were identified in *C. uncialis.* Four terpene synthases possessed accessory genes; three of these are illustrated below. Curiously, one terpene gene was associated with a NRPS-PKS hybrid gene and

a dehydrogenase (Figure 6.6). Although it is possible that two biosynthetic gene clusters are coincidently close together, the proximity suggests that all three biosynthetic units cooperate to produce a single product. This would suggest the presence of a polyketide-peptide-terpene chimeric metabolite. Bioinformatics via BLAST suggests that TERP-1 possesses sequence similarities with squalene or phytoene synthases (Figure 6.15). The second terpene gene designated TERP-2 is possesses a gene characteristic of cytochrome p450 oxidative enzymes (Figure 6.14). BLAST search reveals TERP-2 to likely be squalene synthetase. The third terpene gene cluster possesses two cytochrome p450 oxidative enzymes and type aromatic prenyltransferase (Figure 6.15). Bioinformatics suggest that TERP-3 is characteristic of dimethylallyl tryptophan synthase as previously characterised in the fungus *Aspergillus oryzae*. Three additional terpenes without accessory enzymes were identified. BLAST suggests TERP-4 functions as a squalene-hopene-cyclase gene; TERP-5 closely identified with squalene cyclases, and TERP-6 is characteristic of squalene synthetases. Squalene is a 30-carbon triterpene that is a precursor to the biosynthesis of sterols in plants and animals including cholesterol and vitamin D. Squalene has been suggested to have prophylactic properties against carcinogenesis (Smith & Theresa, 2000; Newmark, 1997). The biosynthetic products of squalene-hopene cyclases, hopenoids, are notable for their physiological role in condensing lipid membranes. Hopenoid biosynthesis is hence thought to be an adaptive response to heat and acidic stressors (Kannenberg & Poralla, 1999; Siedenburg & Jendrossek, 2011).

A seventh terpene gene (TERP-7) was curiously located proximal to a highly reducing PKS, suggesting a possible polyketide-terpene hybrid secondary metabolite within *C. un-cialis* (Figure 6.6).

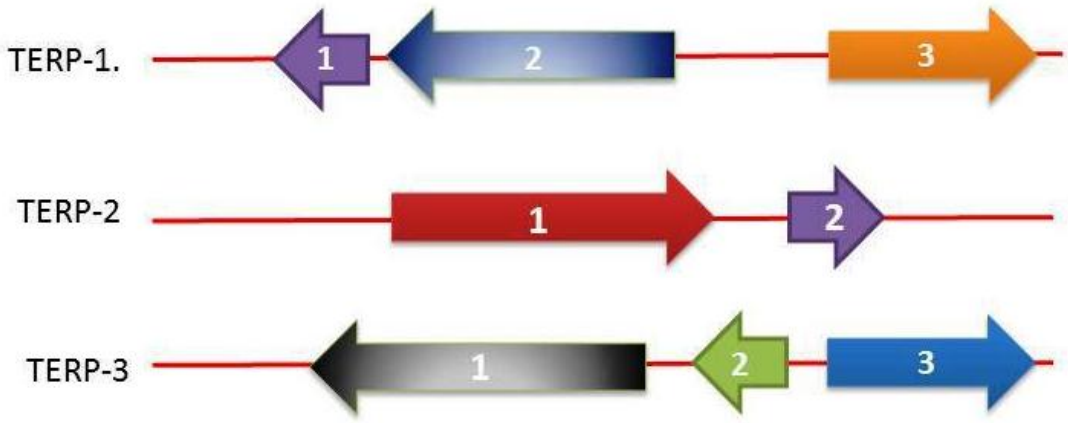


Figure 6.15 Gene clusters of three terpens. TERP-1. (1) TERP-1; (2) dehydrogenase; (3) NRPS/PKS. TERP-2: (1) cytochrome p450; (2) TERP-2. TERP-3: (1) TERP-3; (2,3) cytochrome p450.

6.5 Conclusion

This chapter introduced for the first time the annotation of secondary metabolite gene clusters in *Cladonia uncialis* and one of the first for lichen forming fungi by using antiSMASH based method. Fifty six different gene clusters were identified including, thirty two type 1 different PKS gene clusters; two type III PKS gene clusters; Three independent, novel NRPS biosynthetic gene clusters beside seven noncanonical NRPS genes; four PKS-NRPS hybrid gene clusters; one lanthipeptide synthase gene and seven different terpenes gene clusters. The 32 PKS gene clusters identified in this Chapter include 11 gene clusters of non-reducing PKS genes with no CMet domain, five non-reducing PKS genes with CMet domain and sixteen highly and partially reducing PKS genes. From these results the biosynthetic potential of the lichen *Cladonia uncialis* far exceeds what may be gleaned through traditional surveys of natural products via extraction and isolation. The presence of 56 different gene clusters in *Cladonia uncialis* lichen supports the findings in Chapter 5. The appearance of a large number of secondary metabolites in the mycobiont culture under different conditions (Chapter 5) demonstrates that there is a number of gene clusters in the genome that are silent and only activated under certain conditions. Also the presence of some of these gene clusters transcribed as the anthraquinone gene (Chapter 4) demonstrates that some of these genes might be active.

7 Chapter 7

Protein Analysis and Comparison Between domains of Different *Cladonia uncialis* PKS Genes

Ketosynthase domain

7.1 Introduction

The ketosynthase (KS) domain represents the critical domain in all polyketide synthase genes. This domain is responsible for the process of carbon-carbon bond formation between the growing polyketide chain to give an elongated product via a Claisen condensation reaction. This reaction is a key step in all polyketide and fatty acid syntheses. KSs exist as individual enzymes, which are essential components of type II PKS genes or as a domain found in large multidomain enzymes such as in the type I polyketide synthases (PKSs). The KS sequence is conserved among different species and has been subjected to a large number of phylogenetic analyses by different research groups (Kroken et al.,

2003, Muggia L. *et al*., 2008). The crystal structure of the KS domain has been solved in a type 1 fatty acid synthase and in a modular PKS gene in the context of stand alone KS–AT didomain, including 6-Deoxyerythronolide B Synthase (DEBS) KS– AT3 (Tang et al., 2007) and KS–AT5 (Tang et al., 2006). The KS domain possesses a thiolase fold, which contains an α/ß/α/ß/α fold architecture that forms in total five-layered core with three layers composed of α-helices interspersed by two layers of a β-sheet. The five layers are connecting with each other by loops (Austin.*et al*., 2003). This thiolase fold is common to all KS domains and was first observed in the 1994 structures of a homodimeric yeast KS and then confirmed by a number of crystal structures on different KS domains (Austin, *et al.*, 2003, Whicher, *et al*., 2014). The catalytic active site in the KS is comprised of a cysteine (from a TACSSS motif) along with two histidine (from EAHGTG and KSNIGHT motifs). It has been found that all of the KS domains share the same fold (α/β/α/β/α architecture) and they all use the same catalytic cysteine residue for covalent attachment of substrates and intermediates. Some differences found among different KS domain crystal structures were differences in the extent and structure of the loops located on the opposite side of the active site cavity, the position and identity of the catalytic residue (excluding the universally conserved cysteine) and the different sub-

strate chain-length specificities for CoA-linked or ACP-linked thioesters (Whicher, *et al.*, 2014, Tsai, S., Ames, 2009, Austin, *et al.*, 2003).

Although all the KS domains have the same fold pattern, they are different in their specificity toward different substrates. The specificity of type I and II FAS and PKS KSs are significantly different. In general FAS KS domains are known for their high specificity toward saturated acyl chains while type I modular KS domains have a wide range of substrate specificities that can vary in length from diketide to decaketide. However the substrate specificity of each KS domain is controlled by the size and shape of the KS substrate-binding channel (Whicher, *et al*., 2014, Tsai, *et al.*, 2009, Austin, *et al.*, 2003). This channel is divided into two halves, the substrate-binding region and the Phosphopantetheine (PPT)-binding region. The PPT-binding region is well conserved among different KS domains while the acyl-binding region varies significantly (Tsai, et al., 2009).

Another main domain in all PKS genes and fatty acid synthase genes is the acyltransferase (AT) domain which selects and loads the extender unit on the ACP domain. It is known that the AT domain from fatty acid synthase and iterative PKS genes uses malonyl-CoA as an extender unit while the AT domain from modular PKS genes has much wider substrate specificity toward different substrates (Xu, *et al*., 2013).

Due to the importance of these two domains (KS and AT) and to develop an understanding of the structure and domain specificity toward different substrates of PKS-KS and

PKS-AT domains from lichen fungi, two KS (KS1 and KS2) domains and two AT (AT1 and AT2) domains from two different non-reducing PKS genes from *Cladonia uncialis* have been investigated. The amino acid sequences were confirmed by mass spectrometry and models of the protein fold were generated by two different techniques. Also two AT domains from *Cladonia.uncialis* non reducing PKS genes were investigated. The amino acid sequence from AT1 was partially confirmed by the mass spectrometry. The amino acid sequence used in modeling AT2 domain was translated from the genome nucleotide sequence. KS1 and AT1 were amplified from the proposed halogenated anthraquinone gene (Chapter 4) while KS2 and AT2 were amplified from the non-reducing PKS gene that is proposed to produce methylphloroacetophenone (Chapter 3).

7.2 Material and Methods

7.2.1 Amplification and sequencing, of the DNA

The four domains (KS1 and AT1 from *Cladonia uncialis* anthraquinone gene,Chapter 4 and KS2 from *Cladonia uncialis* usnic acid gene Chapter 3) were PCR amplified (TC-3000) by using the primer sets shown in Table 7.1.

PRIMER NAME	PRIMER SEQUENCE
KS1 sumo F	CGCGAACAGATTGGAGGTTGTACTCACTGCGGTCGTCT
KS1 sumo R	GTGGCGGCCGCTCTATTA GTATGCTCGCCTCTTTCCAC
KS2 sumo F	CGCGAACAGATTGGAGGTGACCCTTTTGGATCGGTTTC
KS2 sumo R	GTGGCGGCCGCTCTATTAATTGATGCAAGCGGCTAATC
AT1 sumo F	CGCGAACAGATTGGAGGTCAGCCTCTCCGTGTCTCCT
AT1 sumo R	GTGGCGGCCGCTCTATTATGGTAGGCCAAGAACCACAT

Table 7.1 Primers used to amplify KS1 and KS2. All the primers have overhang sequence that are complementary to PETite SUMO vector sequence.

The first 18 nucleotides at the 5' end of the primers were designated to be an overhang sequence that match the two ends of the PETite N-His SUMO vector (Lucigen, USA) which was used to express the KS and the AT domains. The following PCR reaction was used: 1X long amplification polymerase buffer, 200 μM of each dNTP, 0.5 μM of each primer, 2 Unit/ μL of long amplification polymerase (New England Biolab, Canada) and water up to 50 μL. The PCR reaction conditions were as follows: initial denaturation at 95 ^{o}C for 2 minutes followed by 30 cycles of 94 ^{o}C for 30 s, 15 s at 56 ^{o}C as annealing temperature for all the primers used and then 65 ^{o}C for 4 minutes as extension. A finial polymerization was carried out at 65^{o}C for 10 minutes and the DNA fragment was then purified from the gel and stored at 4^{o}C.

7.2.2 KS and AT domains protein expression

Almost 25 ng of the PETite vector (see Chapter 2, 2.11.2) (Lucigen, USA) was mixed with 30 ng of the purified DNA fragments separately and then mixed with 40μL of thawed 10 G Chemically Competent Cells (H10 CCC), (Lucigen, U.S.A.) and stirred with a pipet tip. For the vector background control 25 ng of the empty pETite vector

DNA was added into 40 uL of H10 CCC and mixed as above. For the positive control insert a 25 ng of the pETite vector DNA and 50 ng of positive control insert DNA was added to the 40 uL of H10CCC and mixed. The cells and the DNA were transferred into a pre-chilled polypropylene culture tube (Fisher, Canada) and then incubated in ice for 30 minutes. The transformed *E. coli* was then heat shocked at 42 ^{o}C for 45 seconds and then returned back into the ice for more two minutes. Thereafter 960 µL of room temperature Recovery Medium (Lucigen, USA) was added to the cells in the culture tube and then placed in a shaking incubator at 250 rpm for 1 hour at 37^{o}C. Transformed cells (100 µL and 50 µL) were inoculated separately on Luria Bertani (LB) agar plates containing 30 µg/mL of kanamycin by using a glass spreader method. The plates then were incubated at 37 ^{o}C overnight (18 hours). Putative transformants were identified by PCR. Each colony was added into a small PCR tube with the following reagents: 1X long amplification polymerase buffer, 200 µM of each dNTP and 0·5 µM of each primer. The colony was boiled with the reagents at 95 ^{o}C for 10 minutes to break the bacterial wall to release the plasmid. After the PCR tube cooled down, 2 Unit/ µL long amplification polymerase was added (New England Biolab, Canada) and water up to 50 µL. The PCR reaction condition was as follows: initial denaturation at 95 ^{o}C for 2 minutes followed by 30 cycles of 94 ^{o}C for 30 s, 15 s at 56 ^{o}C as annealing temperature and then 65 ^{o}C for 4 minutes as extension. A finial extension was carried out at 65 ^{o}C for 10 minutes. For more confirmation the putative transformants were inoculated into 5 mL of liquid LB medium containing 30 µg/mL kanamycin and placed in a shaking incubator at 250 rpm overnight. Plasmid DNA was harvest from 3 mL of overnight cultures and purified by using (Gene

JETTM Plasmid Miniprep Kit, Fermentas, Canada) and eluted in a final volume of 50 µL of nuclease free water. The sequence of the two plasmid DNAs (PETite N-His SUMO vector/ KS) were confirmed at MICB DNA sequencing services (Manitoba Institute of Cell Biology, Cancer Care Manitoba, University of Manitoba).

The PETite N-His SUMO vector/ KS domains constructs (30 ng) were subsequently transformed into the *E.coli* BL21 (Lucigen, USA) and then incubated with 960 µL of Recovery Medium in the shaker at 250 rpm and 37 ^{o}C for one hour. One to one hundred µL of transformed cells were plated on LB agar plates containing 30 µg/mL kanamycin and then incubated overnight at 37 ^{o}C to allow the bacteria to grow. A few colonies of the growing BL21 cells were mixed with 10 mL LB liquid media that has 30 µg/mL kanamycin solution and incubated in the shaker at 250 rpm and 37 ^{o}C for overnight. 2.5 mL of the growing culture were added to 250 mL of LB media with 30 µg/mL kanamycin solution and kept in the shaker at 250 rpm and 37 ^{o}C for almost 4 hours till the optical density reading of the samples reached between 0.6 to 0.7 at 600 nanometer. To induce expression, IPTG was added into the culture to a final concentration of 1mM. The samples were kept in the shaker for 4 hours at 250 rpm and 37 ^{o}C and then continued overnight at 150 rpm and 19 ^{o}C for over expression. The induced cells were collected and pelleted at 7,000 rpm for 20 minute and stored at -80 ^{o}C.

7.2.3 KS and AT domains protein purification

After expression, the cells were lysed and purified by using B – PER® 6X HIS fusion protein purification kit (Thermo scientific, Canada). The cells were harvested by centrif-

ugation (5000 rpm, 20 minutes, 4 °C), and resuspended in 10 mL B-PER reagent lysis buffer (20 mM Tris buffer, pH 7.5, and unknown proprietary additive) by vortexing. The soluble protein were separated by centrifugation at 14000 rpm for 15 minutes and then loaded into the gravity nickel chelate column. The nickel resin was washed by two different washing buffers, buffer one (35 mM Tris, 150 mM NaCl, 10 mM imidazole, 5% glycerol, pH 7.2, and 0.5X concentration of proprietary B-PER Reagent additive) and buffer two (50 mM Tris, 300 mM NaCl, 25 mM imidazole, 10% glycerol, pH 6.8) and then purified by elution with a buffer that has a high concentration of imidazole (50 mM Tris, 300 mM NaCl, 200 mM imidazole, 10% glycerol, pH 6.8).

The purified SUMO taged proteins were electrophoresed on 12 % sodium dodecyl sulfate-Polyacrylamide gel electrophoresis (SDS-PAGE). Mass spectrometry, matrix-assisted laser desorption/ionization- time-of-flight (MALDI TOF) was applied to confirm the proteins amino acid sequences. All mass spectrum experimental analysis was done by Dr. Lynda Donald, University of Manitoba. Protein modelling for the KS domains were done by Dr. J. Stetefeld, Department of Chemistry, University of Manitoba. Protein homology modeling was performed using the Swiss-Model server. The generated protein models were visualized using PyMOL molecular visualization system, version 1.7.2. AT domains modeling were done by using Swiss-Model server and the generated proteins models were visualized using LASERGENE Protean 3D software (http://www.dnastar.com/t-protean-3D.aspx).

7.3 Results and Discussion

The KS1 and AT1 domains were amplified from the *Cladonia uncialis* the non-reducing PKS gene that has been proposed to produce an anthraquinone (Chapter 4). The second KS domain (KS2) came from the *Cladonia uncialis* non-reducing PKS gene which been proposed to produce methylphloroacetophenone. The two KS domains and the AT domain were overexpressed in *E. coli* . After expression, the cells were lysed and purified by using B – PER® 6X HIS fusion protein purification kit. The SUMO taged purified proteins were electrophoresed on 12 % SDS-PAGE (Figures 7.1, 7.2 and 7.3)

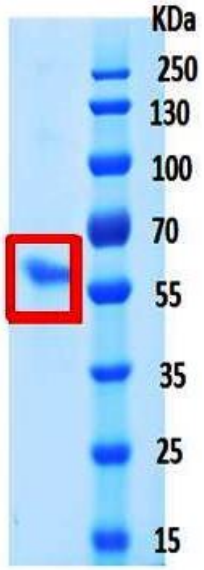


Figure 7.2 SDS-PAGE gel of KS1-SUMO taged purified protein from *Cladonia uncialis* proposed anthraquinone non reducing PKS gene. The red box shows the purified KS1 domain with the SUMO protein at 60 KDa. The molecular weight of the SUMO is 15 KDa and the molecular weight of the KS1 is 45 KDa

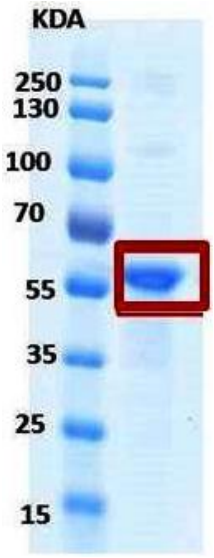


Figure 7.1 SDS-PAGE gel of KS2-SUMO taged purified protein from *Cladonia uncialis* proposed usnic acid non reducing PKS gene. The red box shows the purified KS1 domain with the SUMO protein at 60 KDa. The molecular weight of the SUMO is 15 KDa and the molecular weight of the KS1 is 45 KDa

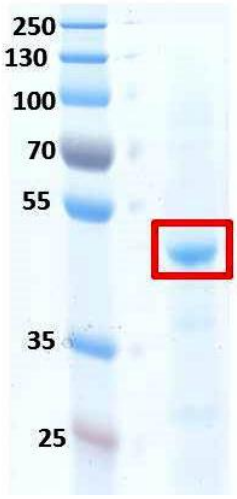


Figure 7.3 SDS gel of AT1-SUMO taged protein from Cladonia uncialis proposed anthraquinone non reducing PKS gene. The red box shows the purified AT1 domain with the SUMO protein at 47 KDa. The molecular weight of the SUMO is 15 KDa and the molecular weight of the AT1 is 32 KDa

Matrix-assisted laser desorption/ionization (MALDI) MS and tandem mass spectrometry analyses were undertaken to confirm the amino acid sequence for the three domains. Proteins were enzymatically digested into smaller peptides using trypsin and introduced into the mass spectrometer and identified by peptide mass fingerprinting by using tandem mass spectrometry. This method consistently revealed the presence of dominant ions as well as smaller ions fractions in both of the domains which matched their amino acid sequences. For KS1 Figures 7.4 and 7.5 shows almost all the amino acid sequences for this domain were confirmed by MALDI MS and tandem mass spectrometry. Underlined peptides have m/z values that match ions seen in the MALDI spectrum, Figure 7.4. Those in red were confirmed by tandem mass spectrometry. The amino acids sequence from 1-107 represents the SUMO sequence (pETite N-HIS SUMO Kan expression vector). Each peak in MALDI spectrum shown in Figure 7.5 is labeled by the amino acids number as

shown in Figure 7.4. For example, the signal observed in the region of m/z 1500 (Figure 7.5) represents the amino acid sequence from 244-256 (VGTFYGQTSDDWR) (Figure 7.4).

001 MHHHHHHGSL QDSEVNQEAK PEVKPEVKPE THINLKVSDG SSEIFFKIKK
051 TTPLRRLMEA FAKRQGKEMD SLTFLYDGIE IQADQTPEDL DMEDNDIIEA
101 HREQIGGSVG DPFGPTKLVK DASGTSGKIA VVGMSGRFPN ADNLESLWSL
151 LEQGLDVHRK VPPDRFDVDA HYDPTGKRKN TSHTPYGCFV KEPGLFDPRF
201 FNMSPREAYQ TDPMGRLALV TAYEALEMSG FMPNRTPSSV LDRVGTFYGQ
251 TSDDWRQVNA AQNIDTYYIP GTIRAFASGR INYYFKFKGP SYNVDTACSS
301 SFSAIQLACT SLQAKECDTV VAGGLNIMTT PDIFSGLSRA QFLSKTGSCR
351 TFDDTADGFC RGDGVATVIL KRLEDAEADN DPILGIILGT ATNHSSEAVS
401 ITQPHAPTQE LLYRKILRNT GVDARDISYV EMHGTGTQAG DGAEMESISN
451 VFAPRVQGRL PGQTVHVGAL KANIGHGEAS AGVASLIKVL LMLQRNAIPP
501 HIGIKGxMNK TFPTDLEERG IRIALQETSW PHPTGGKRRA Y

Figure 7.4 KS1 amino acid sequences. Underlined peptides have m/z values that match ions seen in MALDI spectrum Figure 7.5. Those in red were confirmed by tandem mass spectrometry. Amino acids 1-107 are SUMO. MS analysis done by Dr. Lynda Donald, University of Manitoba.

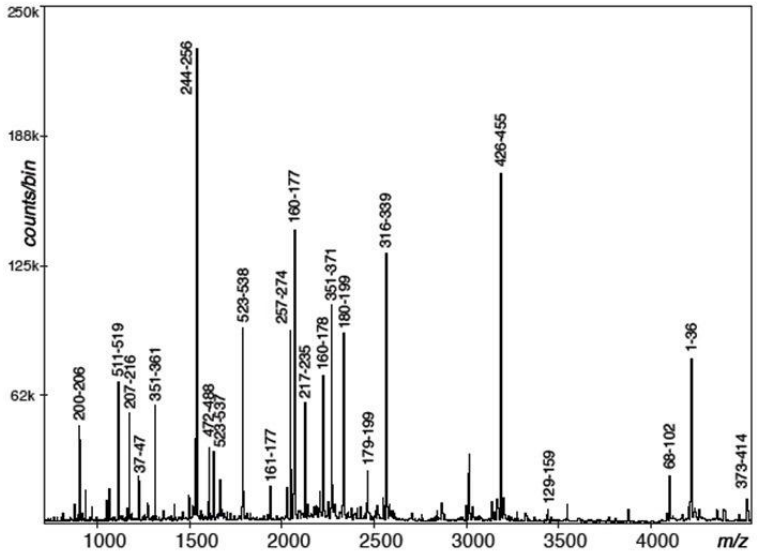


Figure 7.5 MS profile for the KS1. Each peak represent ion that match certain amino acid sequences from KS1. The numbers above each peak represent the position and the number of the amino acid fragment (Figure 7.5). MS analysis done by Dr. Lynda Donald, University of Manitoba.

The majority of the amino acid sequences in the KS2 were also confirmed by (MALDI) MS and tandem mass spectrometry, (Figures 7.6 and 7.7). The amino acid sequences identified in KS2 are not as complete as for KS1 because a large number of the peptides are very large in mass. Since the upper limit for MSMS is about 3500 m/z, a second spectrum was taken over an extended mass range.

001 MHHHHHHGSL QDSEVNQEAK PEVKPEVKPE THINLKVSDG SSEIFFKIKK
051 TTPLRRLMEA FAKRQGKEMD SLTFLYDGIE IQADQTPEDL DMEDNDIIEA
101 HREQIGGDPF GSVSQGYPKD AIAIIGMGCR FPGADSIDEY WNLLTEGKSM
151 LSEIPEARFG RGRPARSNSS LRFWGNFLRD IEAFDHGFFK KSPREAVSMD
201 PQQRVLLQVA YEALESSGYF ADSSRPEDVG CYIGACATDY DFNVASHPPS
251 AYSAIGTLRS FLSGKLSHYF GWSGPSLVLD TACSSSAVAI HTACTALRTG
301 QCSQALAGGI TLMTSPYLYE NFAAAHFLSP TGSSKPFSAD ADGYCRGEGG
351 GLVVLKRLSD ALRDNDHILG VIAGSAVNQN DNCVPITVPH TSSQGNLYER
401 VTEQAGVRPS EVTFVEAHGT GTPVGDPIEM ESIRRVFGGL HRVAPLIVSS
451 AKGNIGHLEG ASGVAALIKA LLQMEHHLAP RQASFKTLNP KIPALEPDNL
501 CIPTSNLA

Figure 7.6 KS2 amino acid sequence. Underlined peptides have m/z values that match ions seen in MALDI spectra (Figure 7.7). Those in red were confirmed by tandem mass spectrometry. Amino acids 1-107 are SUMO. MS analysis was done by Dr. Lynda Donald, University of Manitoba.

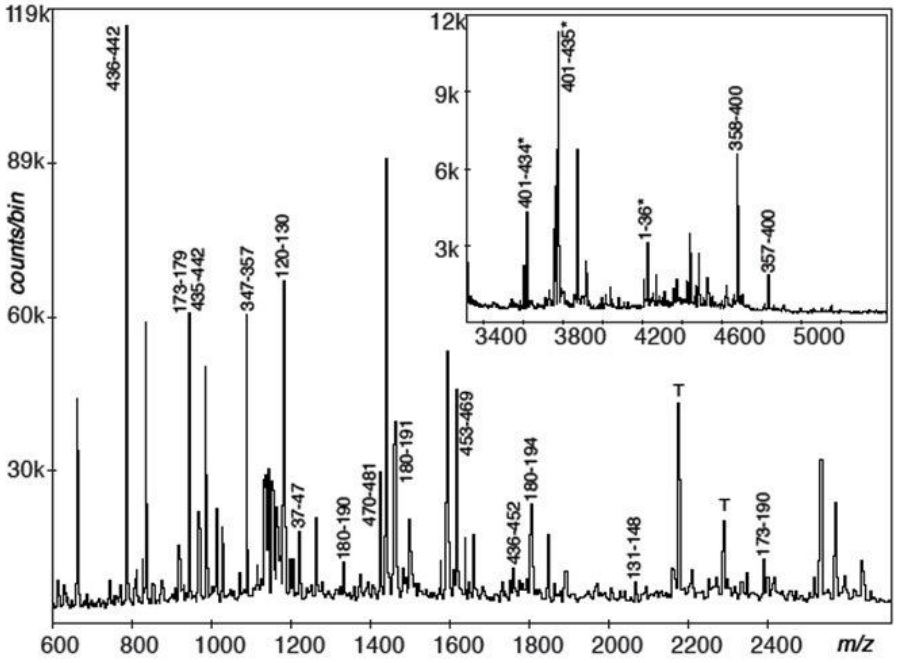


Figure 7.7 MS profile for the KS2. Each peak represents an ion that matches certain amino acid sequences from KS2. The numbers above each peak represent the position and the number of the amino acid fragment (Figure 7.8). The inset spectrum represents the spectrum that was taken at extended range. MS analysis was done by Dr. Lynda Donald, University of Manitoba.

To compare the similarity between the amino acid sequences in the two domains, the sequences were aligned together by using ClustalX 2.0.7. The alignment showed a 30.5 percent sequence identity as shown in Figure 7.8. The two domains have the same active site motif (TACSSS) and they also have the same histidine residue motifs with one amino acid different in the second domain (HGTGT), (KANIGH) in KS1, (HGTGT), (KGNIGH) in KS2, Figure 7.9.

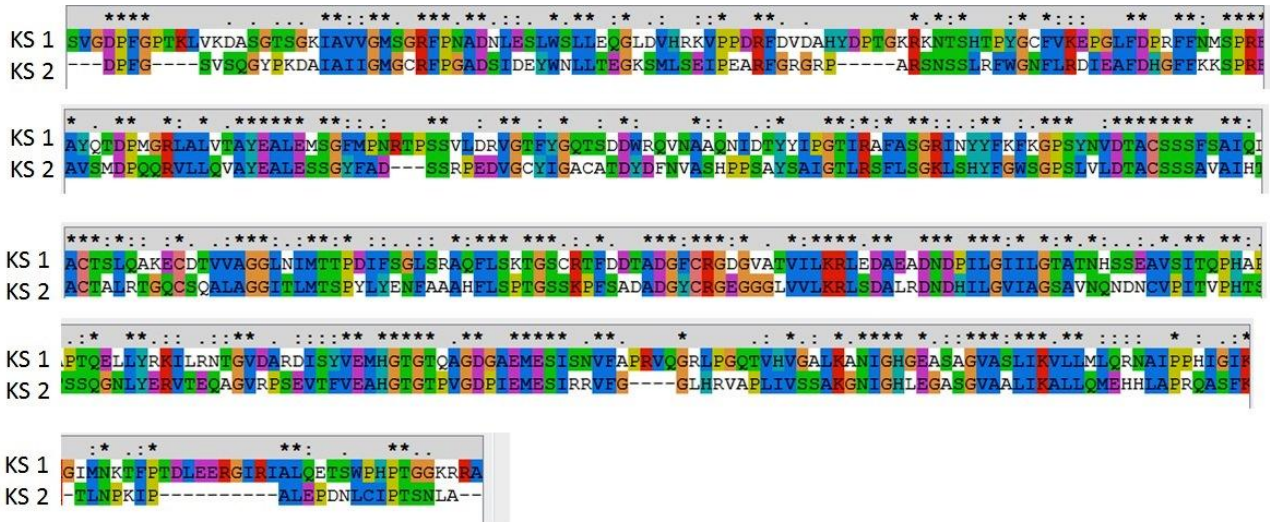


Figure 7.8 Amino acid alignment between KS1 and KS2 by using ClustalX 2.0.7. Asterisks indicate residues exactly conserved between the two domains. Homologous amino acids have the same colour.

The folding patterns for the two domains were investigated to see if the differences in the amino acid sequences would make a difference in their folding pattern. Using PyMOL software, version 1.7.2., homology models were visualized for KS1 and KS2 Figures 7.10, 7.11, 7.12. KS1 domain model showed sequence identity 38% to KS domain from module CurL of the curacin A polyketide synthase while KS2 shows 46% identity to the

same gene (Whicher *et al.*, 2013). Both of the KS domains forms homodimers and each monomer possess a thiolase fold, which contains α/ß/α/ß/α fold architecture. The catalytic active site in each one of them is comprised of a cysteine (TACSSSF) in KS1 and (TASSSA) in KS2 beside two histidines (from EXHGTG and KXNIGH motifs in KS1 and KS2, respectively) as shown in Figure 7.9, in which the cysteine is the nucleophile involved in the transthioesterification reactions. Previous studies showed that the cysteine is hidden in the substrate binding pocket and can only be reached by a phosphopantetheinyl arm attached to the ACP domain (Tang *et al*., 2006).

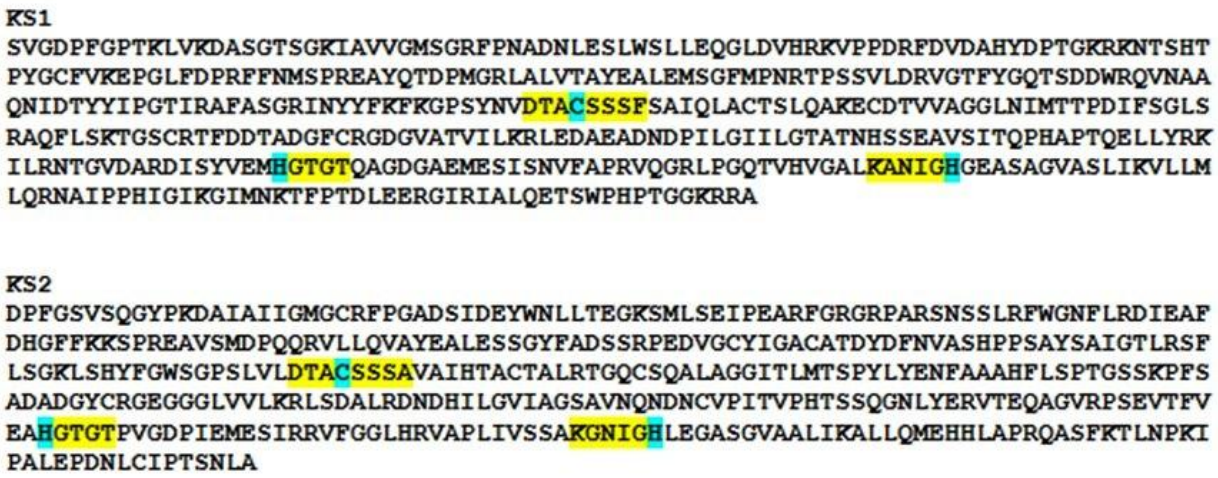


Figure 7.9 Amino acid sequences of the two KS domains with their active sites. The blue colour highlights the active site cysitine and the two histidines and the yellow colour highlights their sequence motifs

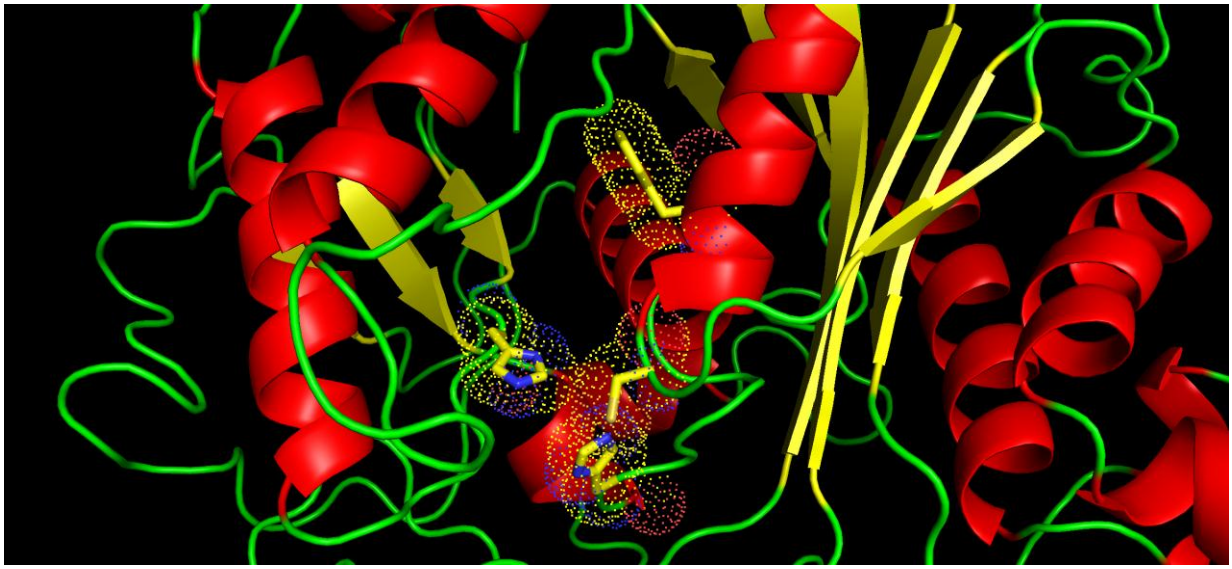


Figure 7.10 Ribbon representation of KS1 model colored by its secondary structure. The active site residues cysteine, two histidine residues and phenylalanine are colored in yellow and shown in stick representation.

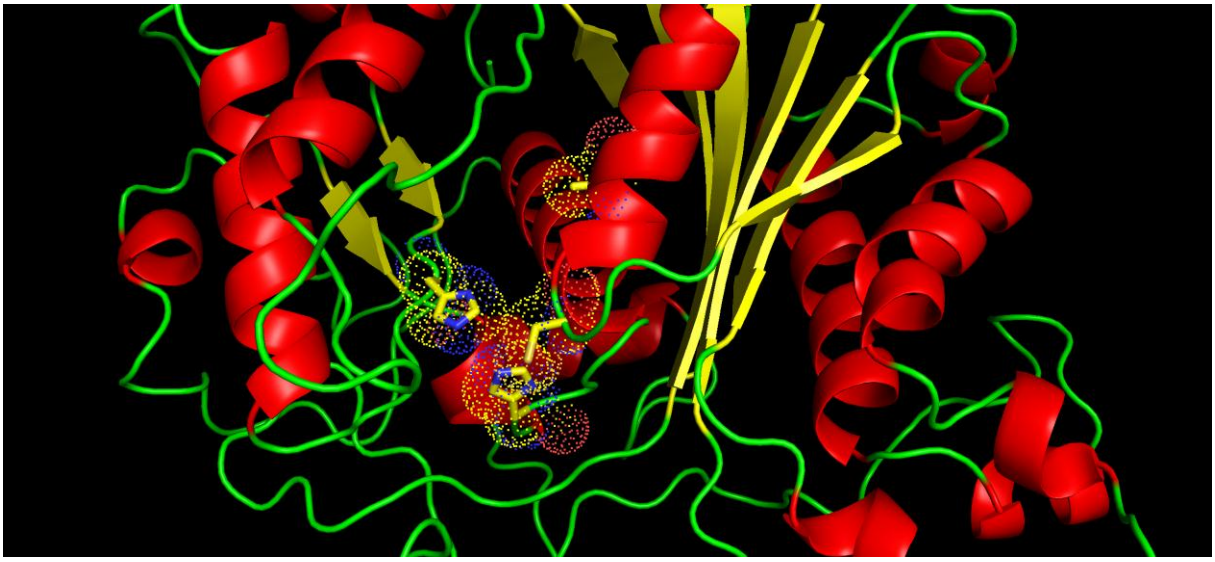


Figure 7.11 Ribbon representation of KS2 model colored by its secondary structure. The active site cysteine, two histidine residues and alanine are colored in yellow and shown in stick representation.

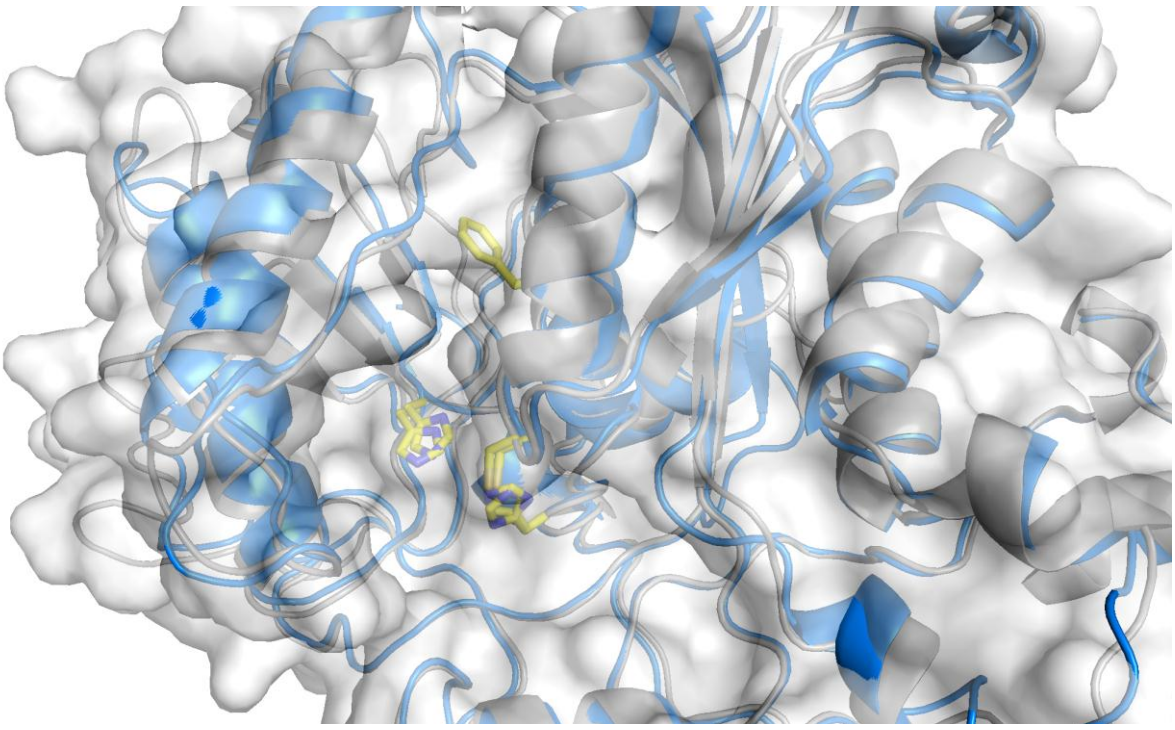


Figure 7.12 Alignment of KS1 and KS2 showing the active site cysteine and histidine residues in each domain and also showing alanine residue in KS1 and phenylalanine in KS2 (active-site residues are shown in yellow)

Different crystal structure studies on the KS domain, showed that the two histidine residues contribute to the formation of an oxyanion hole that stabilizes the tetrahedral intermediate (Xu, W. *et al*., 2013) or to activate the water molecule which in turn attacks the malonyl-ACP to form a carbanion which in turn attacks the acyl and forms the ß-ketoacyl-ACP moiety (Chen, Y. *et al.*, 2011). It has been shown that, the shape and the size of the substrate binding pocket determine the specificity of the KS domain toward the substrate (Xu, W. *et al*., 2013) in addition, the loops located on the opposite side of the active site cavity could be a part of the domain specificity (Austin, MB.et al., 2003, Whicher J., *et al*., 2014, Tsai, S., Ames, B., 2009)

KS1 is thought to be from a non reducing PKS gene which makes the domain responsible for almost 7 rounds of carbon-carbon bond formation. The KS2 from the non reducing PKS gene is responsible for only 3 rounds of carbon-carbon bond formation. The two domains are similar to each other in their folding pattern with some differences. The alignment between the two models was done by using (PyMOL, 1.7.2). The first difference between the two domains was that, the loop located on the opposite side of the active site cysteine is longer in the KS1 than in KS2 as shown in Figure 7.14. The KS2 domain carries out more rounds of carbon-carbon bond formation than the KS1 domain, so that the presence of a longer loop might help the active site to reach into the longer substrate. Also it was noticeable that, there was a bulky amino acid phenylalanine close to the active site cysteine in KS1 while alanine was close to the active site cysteine in KS2. The presence of phenylalanine at this position was noticeable in a number of genes that were thought to produce a series of pigments as shown in Figure (7.13).

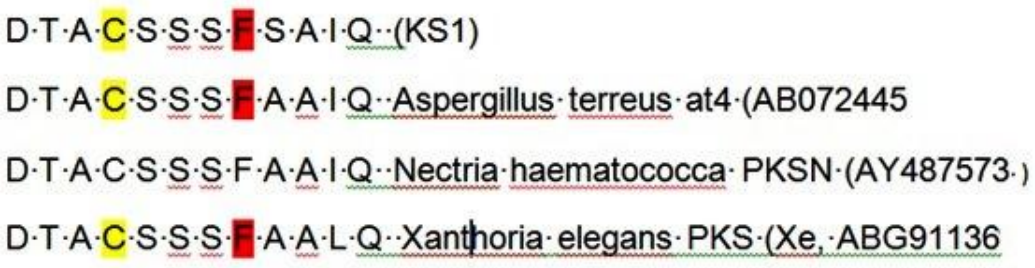


Figure 7.13 KS active site sequence similarity between different non reducing PKS genes that are proposed to produce different kinds of pigments.

Amino acid sequence in the AT1 domain was partially confirmed by tandem mass spectrometry and the confirmed amino acids are in red Figure in 7.14.

Figure 7.14 AT1 domain amino acid sequences. Those in red were confirmed by tandem mass spectrometry. The intensity was not high enough to do MALDI MS to confirm the rest of the sequence.

The AT2 domain nucleotide sequences were converted into amino acid sequences by using the ExPASy translate tool (http://web.expasy.org/translate/). The similarity between the amino acid sequences in the two domains were compared by using ClustalX 2.0.7. The alignment showed that there is 26% percent sequence identity as shown in Figure 7.15. The two domains have the same active site motif (GHSLG) and they also have the same histidine residue motifs with one amino acid different (YAFH), in AT1 and (YGFH) in AT2 (Figure 7.16).

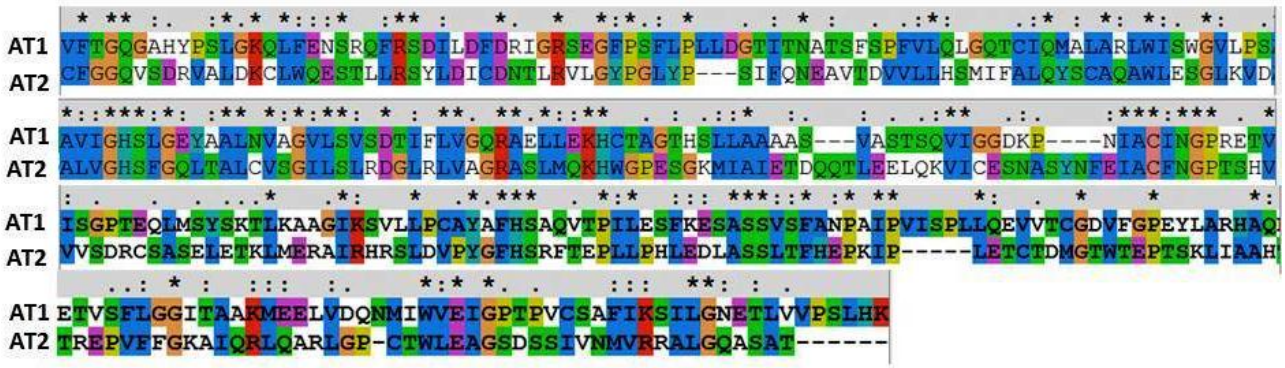


Figure 7.15 Amino acids alignment between AT1 and AT2 by using ClustalX 2.0.7.which shows 26 % identity between the two domains.A sterisks indicate residues exactly conserved between the two domains. Homologous amino acids have the same colour

AT1

VFTGQGAHYPSLGKQLFENSRQFRSDILDFDRIGRSEGFPSFLPLLDGTITNATSFSPFVLQLGQTCIQMALARLWISW
GVLPSAVIGHSLGEYAALNVAGVLSVSDTIFLVGQRAELLEKHCTAGTHSLLAAAASVASTSQVIGGDKPNIACINGPR
ETVISGPTEQLMSYSKTLKAAGIKSVLLPCAYAFHSAQVTPILESFKESASSVSFANPAIPVISPLLQEVVTCGDVFGPEY
LARHAQETVSFLGGITAAKMEELVDQNMIWVEIGPTPVCSAFIKSILGNETLVVPSLHK

AT2

CFGGQVSDRVALDKCLWQESTLLRSYLDICDNTLRVLGYPGLYPSIFQNEAVTDVVLLHSMIFALQYSCAQAWLESGL
KVDALVGHSFGQLTALCVSGILSLRDGLRLVAGRASLMQKHWGPESGKMIAIETDQQTLEELQKVICESNASYNFEIA
CFNGPTSHVVVSDRCSASELETKLMERAIRHRSLDVPYGFHSRFTEPLLPHLEDLASSLTFHEPKIPLETCTDMGTWTE
PTSKLIAAHTREPVFFGKAIQRLQARLGPCTWLEAGSDSSIVNMVRRALGQASAT

Figure 7.16 Amino acid sequences of the two AT domains with their active sites. The blue colour highlights the active site serine and the two histidines while the yellow colour their motifs

The folding patterns for the two domains were investigated by using LASERGENE Protean 3D software (http://www.dnastar.com/t-protean-3D.aspx).

Homology models were created for AT1 and AT2 Figures 7.17, 7.18 and 7.19. AT1 model showed a 30.5 percent sequence identity to AT domain from module CurL of the curacin A polyketide synthase modular PKS gene while AT2 showed 32 percent identity to the same gene (Whicher, *et al.*, 2013)

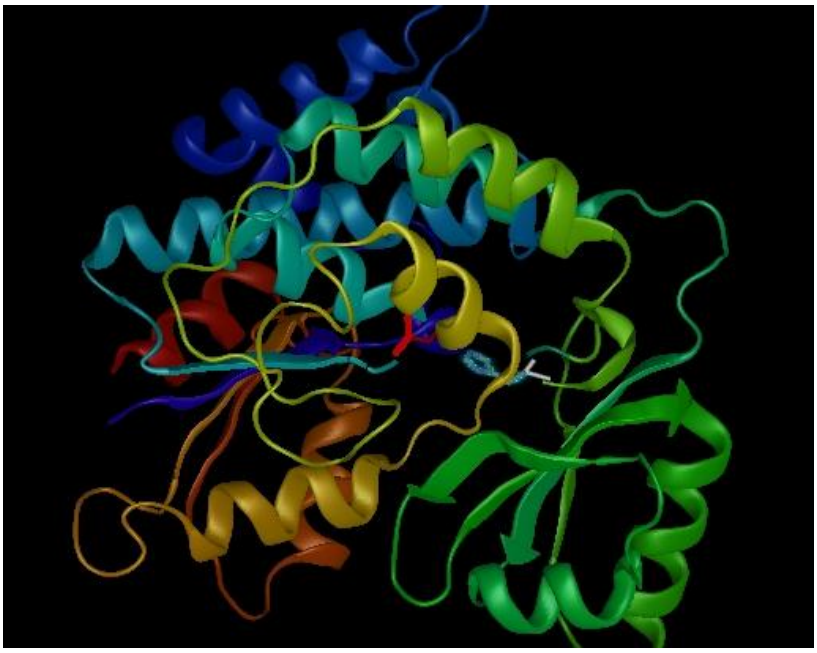


Figure 7.17 AT1 domain (ribbon representation) modeled by using LA-SERGENE Protean 3D software showing serine and histidine active site residues (stick representation).

Figure 7.18 AT2 domain (ribbon representation) modeled by using LA-SERGENE Protean 3D software showing serine and histidine active site residues (stick representation).

Figure 7.19 Alignment between AT1 and AT2 domains by using LA-SERGENE Protean 3D software showing the serine active site residue in each domain (sphere shapes) and the histidine residue in each domain (tube shapes).

Previous crystal structures for the DEBS AT3, AT5, *S. coelicolor* MAT, *E. coli* MAT (FabD) and the porcine FAS AT domain showed that the AT structure has two domains: larger core subdomain which is similar to an α/β-hydrolase fold, with a parallel β-sheet flanked on two sides by α-helices and a smaller subdomain which has a ferredoxin-like structure, consisting of a 4-stranded antiparallel β-sheet capped by two helices (Keatinge-Clay *et al*., 2003; Serre *et al*.,1995; Tang *et al*., 2006, 2007). These results are consistent with the two AT domain models in Figure 7.17 and 7.18 which show the smaller subdomain in green on the right side and the large subdomain which different colours on the left side of the two figures. Using Swiss model software (http://swissmodel.expasy.org/) , the two AT domain modeles showed similarity to other AT domain crystal structures by 30 and 32 percent. The two domains have a small subdomain which consists of 4-stranded antiparallel β-sheet capped by two helices (Figures 7.17 and 7.18). The folding pattern of the large subdomain in AT1 is consistent with the previous crystal structure (Whicher *et al.,* 2013) which is similar to an α/β-hydrolase fold, with a parallel β-sheet flanked on two sides by α-helices (Figure 7.17), while the folding pattern of the large subdomain in AT2 does not show a β-sheet in Figure 7.18. Both of AT1 and AT2 have the same active site motif with one amino acid different: (GHSLG) in AT1 and (GHSFG) in AT2 and they have a histidine residue which is located almost 100 amino acids downstream of the active site serine (Figure 7.16). The active site in each domain is located in the cleft formed between the two subdomains (Figures 7.17 and 7.18). The AT domain catalyzes the transfer of malonyl CoA (in most PKS and fatty acid synthase) into ACP via a ping-pong bi-bi catalytic mechanism using a Ser-His catalytic pair. The nucleo-

philicity of the serine is increased by hydrogen bonding to a conserved histidine. The catalytic reaction of the AT domain starts when the 3-carboxylate of malonyl- or methylmalonyl-CoA forms charge–charge interactions with the side-chain of a highly conserved Arg. Then the active site serine undergoes nucleophilic attack on the thioester carbonyl group, resulting in the formation of a tetrahedral intermediate (Keatinge-Clay et al., 2003; Rangan and Smith, 1997; Serre et al., 1995). The alignment between the two models shown in Figure 7.19 displays the similarity between the two domains in the position of the active site serine, the histidine residue and the folding pattern of the small subdomain, while there is a difference between the large subdomains in the two AT domains. The larg subdomain fold in AT1 is similar to the α/β-hydrolase fold; with a parallel β-sheet flanked on two sides by α-helices (Figure 7.17) while the AT2 large subdomain doesn't have a β-sheet in its folding pattern.

Conclusion

In this Chapter, the investigation of four different domains (KS1, KS2, AT1 and AT2) from different *Cladonia uncialis* non-reducing PKS genes was carried out. Three domains (KS1, KS2 and AT1) were successfully expressed in *E.coli* BL21 and their amino acid sequences were confirmed by using MALDI MS. Modeling of the two KS domains based on their amino acid sequence was performed using the Swiss-Model server. The generated protein models were visualized using PyMOL molecular visualization system. KS1 domain model showed sequence identity 38% to KS domain from module CurL of the curacin A polyketide synthase while KS2 shows 46% identity to the same gene (Whicher *et al.,* 2013). Both domains have the same thiolase folding pattern which contains α/ß/α/ß/α fold architecture. The alignment between the two models showed that they are highly homologous in structure except that KS1 has a longer loop which is located on the opposite side of the active site cysteine. This result is consistent with the fact that the KS1 might do more than 7 rounds of carbon-carbon bond formation and needs a longer loop to enable the active site to reach the substrate. It was also noticeable that there was a phenylalanine amino acid located close to the active site cysteine in KS1 while alanine was present in KS2. The presence of phenylalanine was found among a number of genes

that are thought to produce different pigments. More study needs to be done to explain the effect of each amino acid in the KS specify for choosing the substrate.

Modeling of the two AT domains based on their amino acid sequences was done by using Swiss-Model server. And the generated protein models were visualized using LA-SERGENE Protean 3D software (http://www.dnastar.com/t-protean-3D.aspx). AT1 model showed a 30.5 percent sequence identity to AT domain from module CurL of the curacin A polyketide synthase modular PKS gene while AT2 showed 32 percent identity to the same gene (Whicher, *et al.*, 2013).

The two domains have the same active site motif GHSXG and they have a histidine residue which is located 100 amino acids downstream of the active site serine. Each domain has a small sub domain which has 4-stranded antiparallel β-sheet capped by two helices (Figures 7.17 and 7.18). The AT1 domain has a large sub domain which has a folding pattern similar to the α/β-hydrolase fold, with a parallel β-sheet flanked on two sides by α-helices (Figure 7.17) while the AT2 domain has large sub domain with only α-helices.

8 Chapter 8

Conclusion and Future Work

In this thesis, polyketide biosynthesis in the lichen fungi *Cladonia uncialis* was investigated with the original goal of identifiying the gene cluster responsible for usnic acid production.

The PKS gene that is involved in the biosynthesis of methylphloroacetophenone, the precursor of usnic acid, was studied. The identification of the usnic acid gene cluster was accomplished through *de novo* sequencing of the genome of *Cladonia uncialis*, and in silico analysis of polyketide biosynthetic gene clusters.

The usnic acid polyketide synthase gene cluster has a single PKS gene which contains a ketosynthase domain, an acyltransferase and acyl carrier protein domain, a methyltransferase domain, and no reductive domains, consistent with biosynthesis of the usnic acid precursor, methylphloroacetophenone. Adjacent to the polyketide gene, a single oxidative gene was identified, which is needed for the oxidative dimerization of two molecules of MPA into usnic acid. The transcriptional expression of both the polyketide synthase gene and the accompanying *mpao* gene was confirmed by reverse transcriptase-PCR of the mRNA. This is the first report of the linking of a gene cluster to metabolite production in

lichens and the first identification of a gene cluster responsible for usnic acid biosynthesis. This work will provide the knowledge required for the functional heterologous expression of these genes.

This thesis work also investigated one of the non-reducing PKS genes which appear to be responsible for the production of halogenated lichen metabolites. Halogenated secondary metabolites are an important subclass of natural compounds because of their potent biological activities. The complete understanding of secondary metabolite biosynthesis requires the linking of metabolites to their biosynthetic genes clusters. This is an essential first step in achieving heterologous expression of gene products. The complete sequence of a polyketide synthase (PKS) and accessory genes from the lichen *Cladonia uncialis*, was achieved by *de novo* whole-genome DNA sequencing of the mycobiont as well as cDNA sequencing of the reverse transcribed mRNA. A homology search using BLAST and antiSMASH revealed that the PKS is composed of only the minimal domain components (ketosynthase, acyltransferase, acyl carrier protein), but is flanked by a downstream halogenase and O-methyltransferase as well as an upstream monooxygenase and a reductive gene of unknown function. However, no halogenated secondary metabolites were detected in *C. uncialis* extracts. Based on the biosynthetic operations of the gene cluster, as well as catalogued examples of halogenated polyketides isolated from lichen fungi to date, this study suggests that the gene cluster codes for the biosynthesis of an unidentified halogenated anthraquinone. This is the first report of a halogenated non-reducing PKS gene cluster from lichen fungi.

This thesis also investigated the effect of different growth media on secondary metabolites produced by the lichen fungi *Cladonia uncialis* cultured in the laboratory. The fungal partner was grown on media of varying conditions and the secondary metabolites profile examined by HPLC. A number of secondary metabolites were detected in the fungal culture that were not detected in the field sample. The HPLC results showed that usnic acid was not produced in any of the cultured mycobionts while a range of secondary metabolites were detected in different crude extracts. It was also noticeable that having sucrose in the media stimulates certain metabolites to be produced in high concentrations. Two different secondary metabolites (21.9 minutes in Figure 5.5 and 24.6 minutes in Figure 5.6) were also noticeable in the mycobionts that were grown in trace elements media and sucrose trace elements media respectively. The results of this research represent one strategy for changing the metabolic profile of the organism via changes in the culture conditions.

The number of secondary metabolites detected by HPLC was partly consistent with the number of gene clusters that were detected in the *Cladonia uncialis* fungal genome sequence (56 genes with their accessory genes).

In this thesis also, *de novo* whole-genome sequencing, contig assembly, and annotation of secondary metabolite gene clusters of a species of lichen was reported. The biosynthetic potential of the lichen examined far exceeds what has been reported from extraction and isolation techniques. A total of 56 putative biosynthetic gene clusters were identified within the mycobiont genome of the lichen, *Cladonia uncialis.* Gene clusters and accessory genes conform to major biosynthetic classifications including non-ribosomal peptide

synthetases (NRPSs), polyketide synthases (PKSs), lanthipeptide synthases, terpene synthases, and hybrid genes. Lichen fungal genome sequencing notably revealed a remarkable number and variety of polyketide synthases, including iterative Type I highly reducing, partially reducing, and non-reducing synthases, as well as type III PKSs. To our knowledge, this is the first comprehensive profiling of all secondary metabolite gene clusters of a species of lichen fungus.

Thirty two different type I PKS (non-reducing, partly reducing and highly reducing PKS genes) gene clusters were identified as well as two gene clusters of type III PKS genes. Three independent, novel NRPS biosynthetic gene clusters as well as seven noncanonical NRPS genes that did not contain condensation (C) domains were also identified. Four gene clusters proposed to be responsible for the biosynthesis of PKS-NRP hybrid products were also detected in the genome sequence. One Lanthipeptide synthase gene and seven different terpene gene clusters were also identified in the genome sequence. In total, fifty six different gene clusters with unknown products were identified in *Cladonia uncialis* genome sequence. Functional expressions of all these gene clusters will open the area toward possible new candidates for antibiotics.

Comparison between two KS domains from different non-reducing PKS genes was also reported in this thesis. This part of the research showed that although the two domains are different in their amino acid sequences, they have the same folding pattern plus their active sites are the same. The alignment between the two models showed that they are very much alike in structure except that KS1 has a longer loop which is located on the opposite side of the active site cysteine. This result is consistent with the fact that KS1 does 7

rounds of carbon-carbon bond formation and needs a longer loop to enable the active site to reach the substrate. Another difference between the two KS domains was the presence of phenylalanine located close to the active site cysteine in KS1 while alanine was present in KS2. However more study is needed to be done to clarify the role of these amino acid residues in the active site and the domain specificity for different substrates. Comparison between two AT domains from different non-reducing PKS genes from *Cladonia uncialis* was also reported. The two domains showed 26% sequence identity between each other in the amino acid sequence. They have the same active site motif and they have the same small subdomain folding pattern. The large subdomain folding pattern in AT1 is consistent with other AT crystal structures which have similarity to the α/β-hydrolase fold, with a parallel β-sheet flanked on two sides by α-helices while the AT2 domain has only α-helices in this subdomain.

For the future work, functional expression of usnic acid PKS genes will give potentially new molecules for pharmaceutical and industrial use. Another possible research to undertake in the future is the functional expression of all *Cladonia uncialis* secondary metabolite gene clusters in heterologous hosts that will also give other candidates for medical use. Further investigations are to be done on KS1, KS2, AT1 and AT2 domains.

9 References

1- Abe, I.; Morita, H. *Nat. Prod. Rep*. **2010** *27*, 809–838 | 809.

2- Agelet, A.; Valle`s, J. *J. Ethnophrmacol*. **2001** *77*, 57–70.

3- Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. *J. Mol. Biol.* **1990**, *215*, 403-410.

4- Alvin, A.; Miller, K.; Neilan, B. *Microbiol Res.* **2014** *169*, 483–495.

5- Amnuaykanjanasin, A.; Punya, J.; Paungmoung, P.; Rungrod, A.; Tachaleat, A.; Pongpattanakitshote, S.; Cheevadhanarak, S.; Tanticharoen, M. *Fems Microbiol Lett.* **2005**, *251*, 125-136.

6- Asahina, Y.; Shibata, S. *Chemistry of Lichen Substances*. Lubrecht & Cramer, **1971** New York.

7- August, PR.; Tang, L.; Yoon, YJ.; Ning, S.; Muller, R.; Yu, TW.; Taylor, M.; Hoffmann, D.; Kim, G.; Zhang, XH.; Hutchinson, R.; Floss, HG. *Chem. Biol.* **1998**, *5*, 69-79.

8- Austin, MB.; Noel, JP. *Nat. Prod. Rep.* **2003**, *20*, 79-110.

9- Beiggi, S.; Piercey-Normore, MD. *J. Mol. Evol.* **2007,** *64*, 528-542.

10- Bentley, SD.; *et al. Nature* **2002**,*417*, 141-147.

11- Bessadottir, M.; Egilsson, M.; Einarsdottir, E.; Magnusdottir, H.; Ogmundsdottir H.; *et al*. *PLoS ONE* **2012**, *12*, e51296.

12- Bingle, L.E.; Simpson, T.J.; Lazarus, C.M. *Fungal Genet. Biol.* **1999**, *26*, 209-223.

13- Blasiak, L.C.; Drennan, C.L. *Acc. Chem. Res.* **2009**, *42*, 147-155.

14- Blin, K.; Medema, M.H.; Kazempour, D.; Fischbach, M.A.; Breitling, R.; Takano, E.; Weber, T. *Nucleic Acids Res.* **2013**, *41*, W204-W212.

15- Bochart, J. *Modern Drug Discovery* **1999**, *2*, 22.

16- Bode, H.B.; Bethe, B.; Höfs, A.; Zeeck, A. *ChemBioChem.* **2002**, *3*, 619-627.

17- Boettger, D.; Bergmann, H.; Kuehn, B.; Shelest, E.; Hertweck, C. *ChemBioChem.* **2012**, *13*, 2363 – 2373.

18- Bohlmann, J.; Meyer-Gauen, G.; Croteau, R. *Proc Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4126–4133.

19- Bok, J.W.; Hoffmeister, D.; Maggio-Hall, L.A.; Renato, M.; Glasner, J.D.; Keller, NP. *Chem. Biol.* **2006**, *13*, 31-37.

20- Boustie, J.; Grube, M.; *Plant Genet Resour*, **2007**, 3, 273-287.

21- Brakhage, A.; Schroeckh, V. *Fungal Genet. Biol.* **2011**, *48*,15–22.

22- Brunauer, G.; Muggia, L.; Stocker-Worgotter, E.; Grube, M. *Mycol. Res.* **2009**, *113*, 82-92.

23- Brunauer, G.; Hager, A.; Martin, G.; Turk, R.; Stocker-Wörgötter, E. *Plant Physiol. Biochem.* **2007**, *45*, 146-151.

24- *Brunauer, G.; Muggia, L.; Stocker-Wörgötter, E.; Martin, G. Mycol. Res.* **2009**, *113*, 82–92.

25- Burkin, A.; Kononenko, G. *Biol. Bulletin* **2014**, *41*, 3: 216-222.

26- Campbell, C.D.; Vederas, J.C. *Biopolymers* **2010,** 93: 755: 763.

27- Cansarana, D.; Kahyab, K.; Yurdakulola, E.; Atakolb, O. *Z. Naturforsch* **2006**, *61c*, 773D776.

28- Challis, G.L. *J. Med. Chem* **2008a**, *9*:2618-28.

29- Challis, G.L. *Microbiology* **2008b,** *154*, 1555-1569.

30- Chan, Y.A.; Podevels, A.M.; Kevany, B.M.; Thomas, M.G. *Nat. Prod. Rep.* **2009**, *26*, 90-114.

31- Chen, G.; Wang, H.; Pei, Y. *J Asian Nat. Prod. Res.* **2014,** *16*, 105–122.

32- Chen, Y.; Kelly, E.; Masluk, R.; Nelson, C.; Cantu, D.; Reilly, P. *Protein sci.* **2011**, *20*, 1659-1667.

33- Child, C.J.; Spencer J.B.; Bhogal. P.; Shoolingin-Hordan, P.M. *Biochemistry* **1996,** *35*, 12267-12274.

34- Chooi, Y.H.; David, M.S.; Mery, A.D.; Fujii, I.; Elix, J.A.; Louwhoff, S.; Ann, C.L. *Mycol. Res.* **2008**, *112*, 147-161.

35- Chooi, Y. H.; Tang, Y. J. *Org. Chem*. **2012**, *77*, 9933−9953.

36- Christopher R.; Jamie S.; Jeffrey K. and David S. Comprehensive Natural Products II: Chemistry and Biology, **2010**.

37- Convey, P.; Chown, S.L.; Clarke, A.; Barnes, D.K.A.; Bokhorst, S.; Cummings, V.; Ducklow, H.W.; Frati, F.; Green, T.G.A.; Gordon, S. *Ecol Monogr* **2014**, *84*, 203-244.

38- Corre; Challis *Nat. Prod. Rep.* **2009,** *26*, 977-986.

39- Cortes, J.; Haydock, S.F.; Roberts, G.A.; Bevitt, D.J.; Leadlay, P.F. *Nature* **1990**, *348*, 176-178.

40- Cox, R. J. *Org. Biomol. Chem.* **2007,** *5*, 2010-2206.

41- Crawford, J. M.; Townsend, C. A. *Nat. Rev. Microbiol* **2010,** *8*, 879-889.

42- Crawford, J.M.; Korman, T.P.; Labonte, J.W.; Vagstad, A.L.; Hill, E.A.; Kamari-Bidkorpeh, O.; Tsai, S.C.; Townsend, C.A. *Nature* **2009**, 461, 1139-1143.

43- Crittenden, P.D.; Porter, N. *Trends Biotechnol.* **1991**, *9*, 409-414.

44- Culberson, C.F. Chemical and Botanical Guide to Lichen Products. University of North Carolina Press **1969**, Chapel Hill.

45- Culberson, C. F.; Armaleo, D. *Exp Mycol*. **1992**, 16, 52-63.

46- Culberson, F. *J. Chromatogr* **1972,** *72*, 113–125.

47- Da Costa Silva, J.A.; Bomfim, R.; Estevam, C.; Antoniolli, A.; de Souza Araujo, A.; Thomazzi, S. *Pharma Biol* **2010**, *48*, 745–752.

48- Dayan, F.E.; Romagni J.G. *Pestic. Outlook* **2001,** *12*, 229-232.

49- De Carvalho, A.; Andrade, P.; Silva, H.; Pereira, C.; Figueiredo, C. *Micron* **2005,** 36, 155–161.

50- DePriest, P. *Annu Rev Microbiol* **2004**, *58*, 273–301.

51- Dewick, P. M. *Medicinal Natural Products: A Biosynthetic Approach;* Wiley, Chichester. **2002.**

52- Dewick, P. M. *Medicinal Natural Products: A Biosynthetic Approach;* second edition, Wiley, Chichester, **2006.**

53- Dischinger, J.; Chipalu, S.; Bierbaum G. *Int. J. Med. Microbiol* **2014**, *304*, 51– 62.

54- Eisfeld, K. Physiology and Genetics, Ist Edition, The Mycota XV, T. Anke and D. Weber (Eds.) © Springer-Verlag Berlin Heidelberg, **2009.**

55- Elix, J.A. Biochemistry and secondary metabolites, in: Nash TH (Ed), *Lichen Biology Cambridge University Press, Cambridge*, **1996,** 154-180.

56- Eustáaquio, A.S.; Gust, B.; Luft, T.; Li, S.M.; Chater, K.F.; Heide, L. *Chem. Biol.* **2003**, *10*, 279-288.

57- Fazio, A.; Adler, M.; Bertoni, M.; Maier, M. *Lichenologist* **2012,** *44*, 533–542

58- Ferrer, J.L.; Jez, J.M.; Bowman, M.E.; Dixon, R.A.; Noel, J.P. *Nat Struct Biol.* **1999,** *6*, 775-784.

59- Francolini, I.; Taresco, V.; Crisante, F.; Martinelli, A.; D'Ilario, L.; Piozzi, A. *Int. J. Mol. Sci*. **2013**, *14*, 7356-7369.

60- Friedl, T. and Budel, B. Photobionts. In: Lichen Biology. (ed Nash III. T. H.), **1996**, Cambridge University Press. Cambridge.

61- Frohman, M.A. *et al*., *Proc. Natl. Acad. Sci. USA* **1988,** *85*, 8998–9002.

62- Fujii, I. *J. Antibiot.* **2010**, *63*, 207-218.

63- Funa, N.; Ozawa, H.; Hirata, A.; Horinouchi, S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 6356–6361.

64- Funa, N.; Awakawa, T.; Horinouchi, S. *J. Biol. Chem*. **2007**, *282*, 14476-14481.

65- Gauslaa, Y.; McEvoy, M. *Basic Appl. Ecol.* **2005**, 6, 75-82.

66- Gerke, J; Bayram, O; Feussner, K; Landesfeind, M; Shelest, E; Feussner, I; Braus, G.H. *Appl environ microb* **2012,** *78*, 8234 – 8244

67- Gevers, W.; Kleinkauf, H.; Lipmann, F. *Proc.Nut. Acad. Sci. U. S.* **1968,** *60*,269

68- Gessler, N.N.; Egorova, A.S.; Belozerskaya, T.A. *Appl. Biochem. Microbiol.* **2013**, 49, 85-99.

69- Gramajo, H.C.; White, J.; Hutchinson, C.R.; Bibb, M.J. *J Bacteriol.* **1991,** *173*, 6475-6483.

70- Gross, H. *Appl. Microbiol. Biotechnol.* **2007,** *75*, 267-277.

71- Grube, M.; Blaha, J. *Mycol. Res.* **2003**, 107, 1419-1426.

72- Grube, M.; DePriest, P.T.; Gargas, A.; Hafellner, J. *Mycol. Res*. **1995**, 99, 1321–1324.

73- Hager, A.; Brunauer, G.; Türk, R.; Stocker-Wörgötter, E. J. *J Chem. Ecol*. **2008**, *34*, 113-120.

74- Hamada, N.; Miyagawa, H. *Lichenologist* **1995**, *27*, 201-205.

75- Hamada, N. *Bryologist* **1993**, 96: 569-572.

76- Hamada, N. *Bryologist* **1996**, *99*, 68-70.

77- Hamada, N.; Miyagawa, H.; Miyawaki, H.; Inoue, M. *Bryologist* **1996**, *99*, 71-74.

78- Hao, X. Y.; Wang, M.; Wang, L.; Wei, J. C. *BMC Genomics* **2014**, *15*, 34.

79- Hauck, M.; Willenbruch, K.; Leuschner, C. *J. Chem. Ecol.* **2009**, *35*, 71–73.

80- Hauck, M.; de Bruyn, U.; Javkhlan, D. *J. Plant Ecol.* **2014**, *7*, 287–297.

81- Hawksworth, D.L.; Hill, D.J. *The Lichen-Forming Fungi*, Blackie, Glasgow and London, **1984**,158 pp.

82- Hawranik, D.; Anderson, K.; Simmonds, R.; Sorensen, J. *Bioorg. Med. Chem. Lett*. **2009,** *19*, 2383-2385.

83- Helfrich, N.; Reiter, S.; Piel, J. *Curr. Opin. Biotechnol.* **2014**, *29*, 107-115.

84- Hertweck, C. *Angew. Chem. Int. Ed*. **2009**, *48*, 4688 – 4716

85- Honegger, R. *Lichen Biology* **1996**, 24–36.

86- Honegger, R. *Bryologist* **2000**, *103*, 307-313.

87- Huneck, S. *New Results on the Chemistry of Lichen Substances*. Springer, New York, **2001.**

88- Huneck, S.; Yoshimura, I. *Identification of Lichen Substances. Springer, Berlin, Heidelberg*, New York, **1996.**

89- Huneck, S. *Naturwissenschaften* **1999**, *86*, 559-570.

90- Huneck, S.; Fortschr *Chem. Org. Naturst*. **2001**, *81*, 1-276.

91- Huneck, S. *Naturwissenschaften* **1999**, *86*, 559-570.

92- Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Omura, S. *Nat. Biotechnol.* **2003**, *21*, 526-531.

93- Ingólfsdóttir, K. *Phytochemistry* **2002,** *61*,729–736.

94- Jensen P. R.; Chavarria, L. K; Fenical, W.; Moore, B.S.; Ziemert, N. *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 203-209.

95- Gershenzon, J.; Dudareva, N. *Nat Chem Biol.* **2007**, *3*, 408 – 414.

96- Jones, P.M.; George, A. M. *Cell Mol. Life Sci.* **2004**, *61*, 682-699.

97- Kannenberg, E.; Poralla, K. *Naturwissenschaften* **1999**, 86: 168–176.

98- Keatinge-Clay, A. T. *Nat. Prod. Rep.* **2012**, *29*, 1050-1073.

99- Keller, N. P.; Hohn, T. M. *Fungal Genet Biol*. **1997**, *21*, 17-29.

100- Khosla et al. *Annu. Rev. Biochem.* **2007**, *76*, 195-221.

101- Kim, J. *Mycologia* **2012**, *104*, 362–370.

102- Koehn, F.E.; Carter, G.T. *Nat. Rev. Drug Discov.* **2005**, *4*, 206-220.

103- Kroken, S.; Glass, N. L.; Taylor, J. W.; Yoder, O. C.; Turgeon, B. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15670-15675.

104- Li C.; Florova G.; Akopiants K.; Reynolds KA. *Microbiology* **2004**, 150, 3463-3472.

105- Liu H.; Reynolds KA. *Metab. Eng*. **2001**, 3, 40-48.

106- Lucia M.; Martin G. *fungal biology* **2010**, *114*, 379–385.

107- Maier, T.; Jenni, S.; Ban, N. *Science* **2006**, *311*, 1258-1262.

108- Manenti, R. *Biodivers Conserv* **2014**, *23*, 1879–1893.

109- Mathey, A.; Spiteller, P.; Steglich, W. *Z Naturforsch c.* **2002**, 57 (7-8), 565-7.

110- Mayer, M.; O'Neill, M. A.; Murray, K. E. *anti-cancer Drugs* **2005**, *16*, 805-809.

111- Medema, M.H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M.A.; Weber, T.; Takano, E.; Breitling, R. *Nucleic Acids Res.* **2011**, *39*, W339-W346.

112- Meslet-Cladière, L.; Delage, L.; Leroux, C.; Goulitquer, S.; Leblanc, C.; Creis, E.; Ar Gall, E.; Stiger-Pouvreau, V.; Czjzek, M.; Potin, P. *Plant Cell* **2013**, *25*, 3089–3103.

113- Miao, V.; Coeffet-LeGal, M. F.; Brown, D.; Sinnemann, S.; Donaldson, G.; Davies, J., 2001.*Trends Biotechnol.* **2001**, *19*, 349-355.

114- Müller, K. Appl. *Microbiol. Biotechnol.* **2001**, *56*, 9-16.

115- Molnar, K.; Farkas, E. *Z Naturforsch c*. **2010**, *C 65*, 157–173.

116-Moore, B.S.; Piel, J. *ANTON. LEEUW. INT. J. G.* **2000**, *78*, 391-398.

117-Muggia, L.; Schmitt, I.; Grube, M. *Mycol. Res*. **2008**, *112*, 277-288.

118- Muller K. *Appl. Microbiol. Biotechnol.* **2001**, *56*, 9-16.

119-Murphy, C.D. *Nat. Prod. Rep.* **2006**, *23*, 147-152.

120-Nakano, C.; Ozawa, H.; Akanuma, G.; Funa, N.; Horinouchi, S. *J of Bacteriol.*, **2009**, *191* (15), 4916-4923.

121-Nett, M.; Ikeda, H.; Moore, B. S. *Nat. Prod. Rep.* **2009**, *26*, 1362-1384.

122-Neumann, C.S.; Fujimori, D.G.; Walsh, C.T. *Chem. Biol.* **2008**, *15*, 99-109.

123-Newman, D.J.; Cragg, G. M. *J. Nat. Prod.* **2012**, *75*, 311-335.

124-Newman, D.J.; Cragg, G.M.; Snader, K.M. *J. Nat. Prod* **2003**, *66*, 1022-1037.

125-Nicholson, T. P.; Rudd, M.; Dawson, M.; Lazarus, C. M.; Simpson, T. J.; Cox, R. J. *Chem. Biol.* **2001**, *8*, 157-178.

126- Okuyama, E.; Hossain, F.; Yamazaki, M. *ShoyakugakuZ asshi* **1991**, *45*, 159-162.

127- Oliynyk, M.; Samborskyy, M.; Lester, J. B.; Mironenko, T.; Scott, N.; *et al*. *Nat. Biotechnol.* **2007**, *25*, 447- 453.

128-Omura, S.; Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Takahashi, C.; Shinose, M.; Takahashi, Y.; Horikawa, H.; Nakazawa, H.; Osonoe, T.; Kikuchi, H.; Shibai, T.; Sakaki, Y.; Hattori, M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12215-12220.

129-Paulsen *et al*. *Nat. Biotechnol.* **2005**, *23*, 873-878.

130-Pel, H. J.; de Winde, J. H.; Archer, D. B.; Dyer, P. S.; Hofmann, G.; *et al. Nat. Biotechnol.* **2007**, *25*, 221-231.

131-Peres, V.; Nagem, T.J.; de Oliveira, F. F. *Phytochemistry* **2000**, *55*, 683-710.

132- Piercey-Normore, M.D.; Depriest, P.T. *Am. J. Bot.* **2001**, *88*, 1490-1498.

133- Proksa, B.; Adamcova, J.; Sturdikova, M.; Fuska, J. *Pharmazie* **1994**, *49*, 282.

134- Ramos, D.; da Silva1, P. *Pharm Biol*. **2010**, *48*(3), 260–263.

135- Rangan, V.S.; Smith, S. J. *Biol. Chem*. **1997**, *272*, 11975–11978

136- Schaefer, B.C. *Anal. Biochem.***1995**, *227*, 255–73.

137- Schmitt, I.; Kautz, S.; Lumbsch, H.T. *Mycol. Res.* **2008**, *112*, 289-296.

138- Schnell, N.; Entian, K.D.; Schneider, U.; Götz, F.; Zähner, H.; Kellner, R.; Jung, G. *Nature* **1988**, *333*, 276–278.

139- Schoenafinger, G.; Marahiel, M. A. *Wiley Encyclopedia Chem. Biol.* **2009**, *3*, 432-440.

140- Schwarzer, D.; Finking, R.; Marahiel, M. A. *Nat. Prod. Rep.* **2003**, *20*, 275-287.

141- Shen, B. *Curr. Opin. Chem. Biol.* **2003**, 7, 285-295.

142- Shibata, S.; Ukita, T.; Tamura, T.; Miura, Y. *Jap Med J*. **1948**, *1*, 152–155.

143- Shrestha, G.; Clair, L. *Phytochem Rev* **2013**, *12*, 229–244.

144- Shukla, V.; Joshi, G. P.; Rawat, M. S. M. *Phytochem. Rev.* **2010**, *9*, 303-314.

145- Shukla, P.; Upreti, DK.; Nayaka, S.; Tiwari, P. *Indian j. tradit. know*. **2014**, *13*, 195-201.

146- Siedenburg, G.; Jendrossek, D. *Appl environ microb* **2011**, *77*, 3905–3915.

147- Smith, S.; Tsai, S.C. *Nat. Prod. Rep.* **2007**, *24*, 1041-1072.

148- Søchting, U. *Lichenologic* **1997**, *68*, 135–144.

149- Sokolov, D.; Zarubaev, V.; Shtro, A.; Polovinka, M.; Luzina, O.; Komarova, N.; Salakhutdinov, N.; Kiselev, O. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 7060–7064.

150- Solhaug, K.A.; Gauslaa, Y.; Nybakken, L.; Bilger, W. *New Phylol.* **2003**, *158*, 91-100.

151- Sokolov, D. N.; Luzina, O. A.; Salakhutdinov, N. F. *Usp. Khim.* **2012,** *81*, 8, 747–768.

152- Spurgeon, S. L.; Porter, J. W. Eds.; Wiley-Interscience: New York, **1981**, *1*, 1-46.

153- Staunton, J.; Weissman, K.J. *Nat. Prod. Rep.* **2001**, *18*, 380-416.

154- Stocker-Woergoetter, E. *Nat. Prod. Rep*. **2008**, *25*, 188-200.

155- Stocker-Worgotter E.; Elix, J.A. *Lichenologist* **2002**, *34*, 351-359.

156-, Strieker, M.; Tanovic, A.; Marahiel, M. *Curr. Opin. Struct. Biol*. **2010**, 20:234–240.

157- Tabrez Khan, S.; Musarrat, J.; Alkhedhairy, A. *Ann Microbiol.* **2014**, *64*, 199–207.

158- Taguchi, H.; Sankawa, U.; Shibata, S. *Chem. Pharm. Bull.* **1969**, *17*, 2054-2060.

159- Taguchi, H.; Sankawa, U.; Shibata, S. *Tetrahedron Lett.* **1966,** 5211-14.

160- Tang, G. L.; Cheng, Y. Q.; Shen, B. *J. Nat. Prod.* **2006**, *69*, 387–393.

161- Tang, G. L.; Cheng, Y. Q.; Shen, B. *J. Biol. Chem*. **2007**, *282*, 20273–20282.

162- Thomas, R. C*hembiochem* **2001**, *2*, 612-627.

163- Timsina, B., Hausner, G., and Piercey-Normore, M. D. *Fungal Biology* **2014**, 118: 896-909.

164- Tsai, S.C.; Ames, B.D. *Methods Enzymol.* **2009**, 459, 17-47.

165- Udwary, D. W.; Zeigler, L.; Asolkar, R. N.; Singan, V.; Lapidus, A.; Fenical, W.; Jensen, P. R.; Moore, B. S. *Salinispora tropica. Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 10376-10381.

166- Van den Berg, M. A.; Albang, R.; Albermann, K.; Badger, J. H.; Daran, J. M. *Nat. Biotechnol.*, **2007**, *26*, 1161-1168.

167- Varga, J.; Rigó, K.; Kocsubé, Farkas, B.; Pál, K. *Res. Microbiol*. **2003**, *154*, 593-600.

168- VARITA, K. O. *Antibiotics in lichens* **1973**, 547-561.

169- Vijayakumar, S.; Viswanathan, S.; Reddy, K.; Parvathavarthini, S.; Kundu, B.; Sukumar, E. *Fitoterapia* **2000**, *71*, 564–566.

170- Wagner, C.; Omari, M. E.; Konig, G. M. *J. Nat. Prod*. **2009**, *72*, 540-553.

171- Walsh, C. T. *Acc. Chem. Res.* **2008**, *41*, 4-10.

172- Wang, Y.; Kim, J. A.; Cheong, Y. H.; Koh, Y. J.; Hur, J. S. *Mycol. Progress* **2011**, *11*, 75-83.

173- Wang, Y. Y.; Liu, B.; Zhang, X. Y.; Zhou, Q. M.; Zhang, T.; Li1, H.; Yu, Y. F.; Zhang, X. L. *BMC Genomics* **2014**, *15*, 34.

174- Wang, Y.; Wang, J.; Cheong, Y.; Hur, J. *Mucobiology* **2014**, *42* (1), 34-40.

175- Wasil, Z.; Pahirulzaman, K.A.K.; Butts, C.; Simpson, T.J.; Lazarus, C.M.; Cox, R.J. *Chem. Sci.* **2013**, *4*, 3845-3856.

176- Webster S. P.; Alexeev D.; Campopiano D. J.; Watt R. M.; Alexeeva M.; Sawyer L.; Baxter R. L. *Biochemistry* **2000**, 39: 516-528.

177- Whicher J.; Dutta, S.; Hansen, D.; Hale, W.; Chemler, J.; Dosey, A.; et al. *Nature*, **2014**, *510*, (560-564).

178- Whittet, R.; Ellis, C. *J Biological Conservation* **2013**, *168*, 19–23.

179- White, T. J.; Bruns, T.; Lee, S.; Taylor, J. *In: PCR Protocols: a guide to methods and applications. (*Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds*). Academic Press, New York, USA,* **1990**, 315–322.

180- Wilkinson, B.; Micklefield, J. *Nat. Chem. Biol.* **2007**, *3*, 379-386.

181- Willey, J. M.; van der Donk, W. A. *Annu. Rev. Microbiol.* **2007**, *61*, 477-501.

182- Whicher, J.*et al. Chem.Biol*. **2013,** *20*, 1340–1351.

183- Xu, W.; Qiao, K.; Tang, Y. *Crit. Rev. Biochem. Mol. Biol.* **2013**, *48*, 98-122.

184- Yeku, O.; Frohman, M.A. *Methods Mol. Biol.* **2011**, *703*, 107-122.

185- Yoshimura, I.; Yamamoto, Y.; Nakano, T.; Finnie, J., In: *Protocols in lichenology. Culturing, biochemistry, ecophysiology and use in biomonitoring*; Kranner I, Beckett RP, Varma AK, Eds; Springer-Verlag: Berlin, **2002**, Chapter 1, pp. 3-33.

186- Yoshimura, I.; Kurokawa, T.; Kinoshita, Y.; Miyawaki, H. *J Hattori La.* **1994**, *76*, 249-261.

187- Yu, N. H.; Kim, J. A.; Jeong, M. H.; Cheong, Y. H.; Jung, J. S.; Hur, J. S., *Mycol. Progress*, **2013**, *12*, 519-524.

188- Zambare, V.; Christopher, L. *Pharmaceutical Biology* **2012**, *50* (6), 778–798.

189- Zazopoulos, E.; Huang, K.; Staffa, A.; Liu W.; Bachmann, B.O.; Nonaka, K.; Ahlert, J.; Cox, R.J. *Org. Biomol. Chem.* **2007**, *5*, 2010-2206.

190- Zazopoulos, E.; Huang, K.; Staffa, A.; Liu W.; Bachmann, B.O.; Nonaka, K.; Ahlert, J.; Thorson, J.S.; Shen, B.; Farnet, C.M. *Nat. Biotechnol.* **2003**, *21*, 187-190.

191- Ziemert, N.; Lechner, A.; Wietz, M.; Millán-Aquiñaga, N.; Chavarria, K. L.; Jensen, P. R. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, E1130-9.

192- Zocher, B. E.; Stocker-Worgotter, E. *Folia Cryptogam Estonica*. **2005**, *41*, 123-134.

193- Yeku, O.; Frohman, M.A. *Methods Mol. Biol.* **2011**, 703, 107-122.

Appendix A: High Resolution Images of Figures 3.8 and 3.9

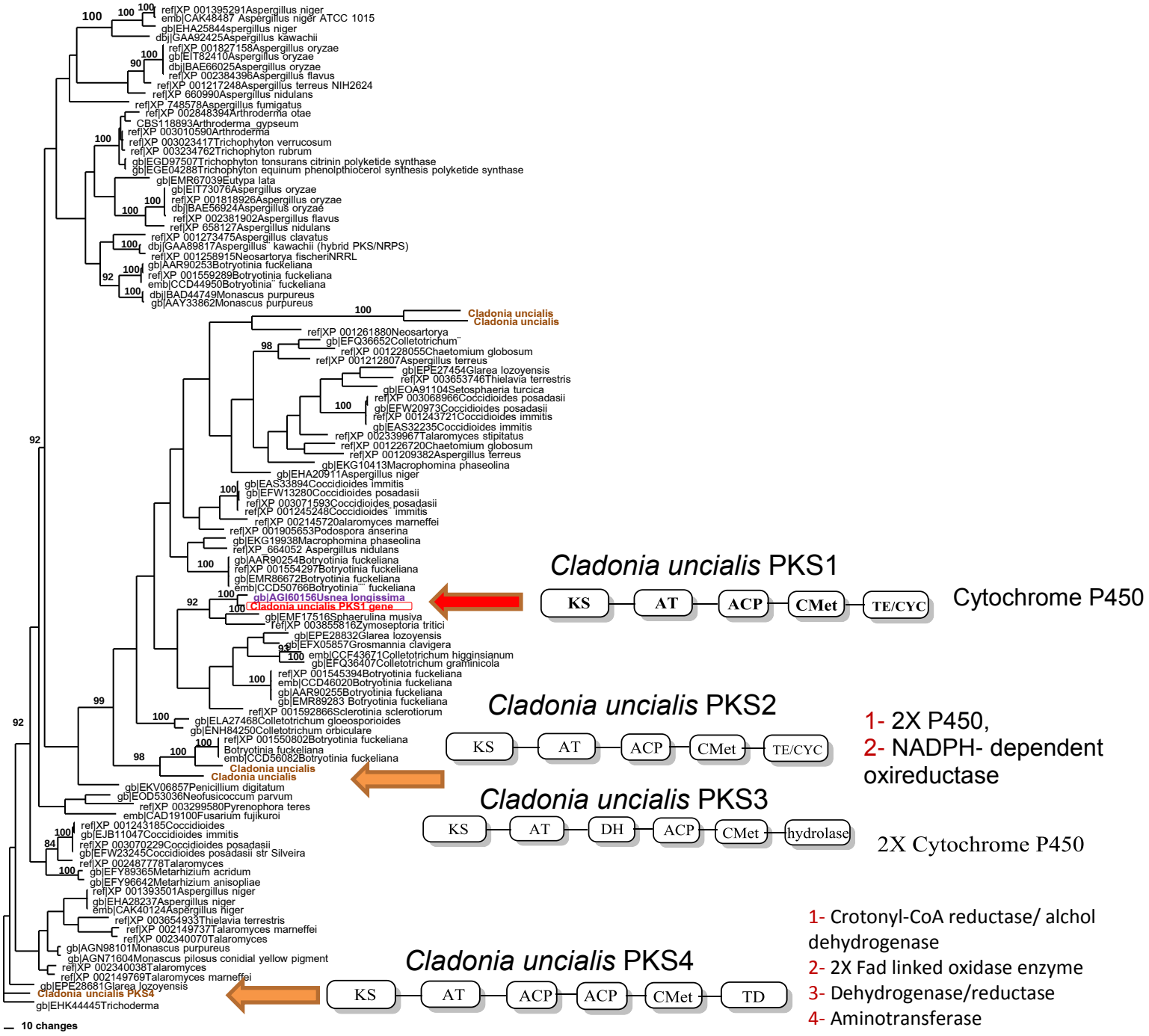


Figure 3.8 Phylogeny of the KS domains from predicted *Cladonia uncialis* usnic acid PKS gene and homologous fungal non-reducing type I PKSs.

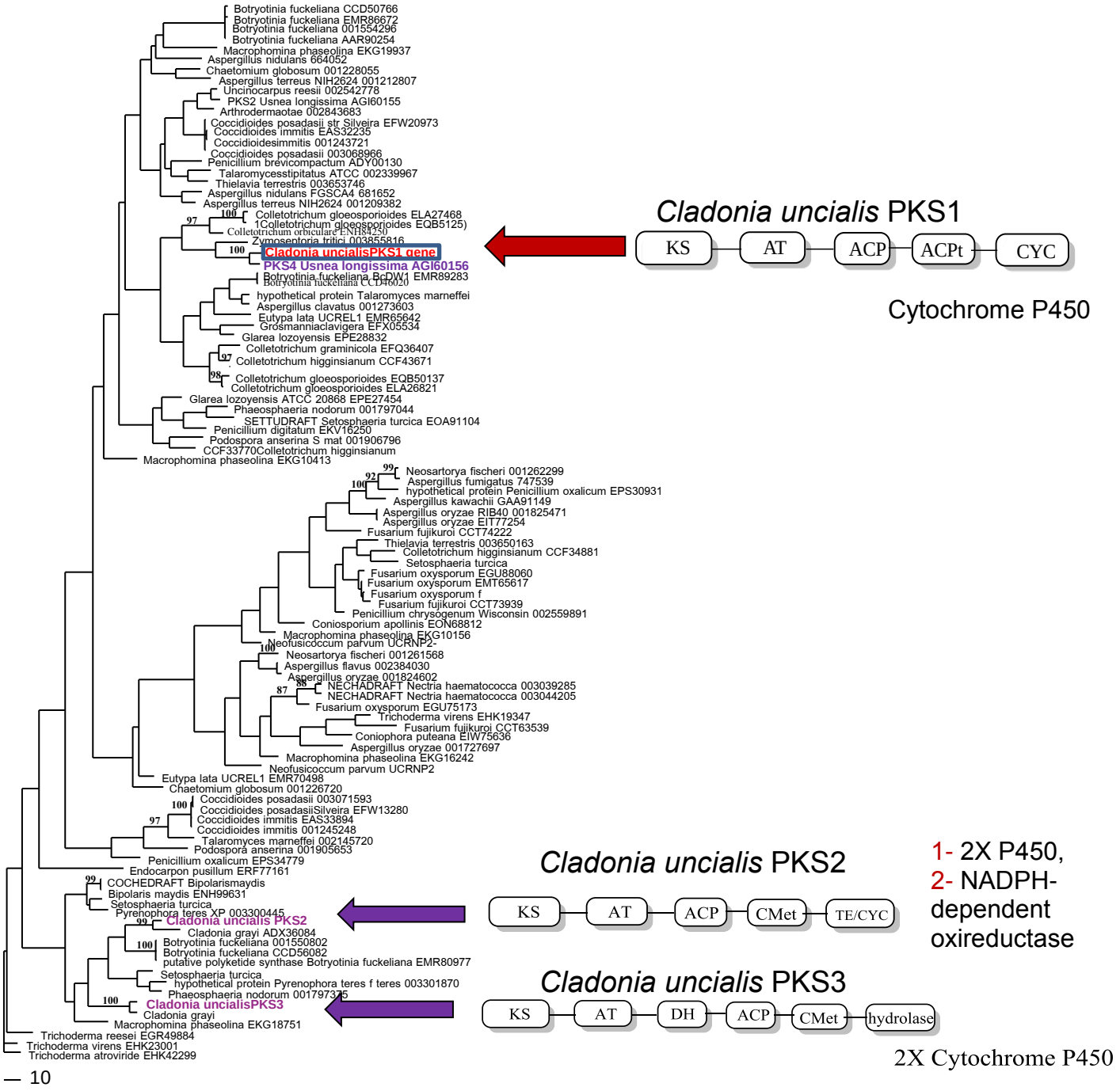


Figure 3.9 Phylogeny of the CYC domains from predicted *Cladonia uncialis* usnic acid PKS gene and homologous fungal non-reducing type I PKSs