THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Systems Biology of the Secondary Metabolism in Filamentous Fungi

Jens Christian Nielsen

CHALMERS

Department of Biology and Biological Engineering
Chalmers University of Technology
Göteborg, Sweden 2018

Systems Biology of the Secondary Metabolism in Filamentous Fungi
Jens Christian Nielsen
Gothenburg, Sweden 2018
ISBN 978-91-7597-671-6

Doktorsavhandling vid Chalmers tekniska högskola
Ny serie nr 4352
ISSN 0346-718X

Department of Biology and Biological Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone +46(0)31-772 1000

Printed by Chalmers Reproservice
Gothenburg, Sweden 2018

Systems Biology of the Secondary Metabolism in Filamentous Fungi

Jens Christian Nielsen

Department of Biology and Biological Engineering

Chalmers University of Technology

Abstract

Filamentous fungi constitute a rich reservoir of pharmaceutically relevant bioactive small molecules. These compounds, commonly referred to as secondary metabolites, are widely used as antibiotics for the treatment of microbial infections, but also as other pharmaceuticals such as immunosuppressors, cholesterol lowering agents and anticancer drugs. Although fungal derived antibiotics have been known for almost a century, genome sequencing has revealed that the biosynthetic potential of fungi is not fully exhausted.

The *Penicillium* genus consists of around 350 accepted species, and many of these are well-known producers of pharmaceuticals and industrially exploited for this. The genus as a whole, however, is grossly understudied at the genomic level. To assess the potential for secondary metabolite biosynthesis in the *Penicillium* genus, we sequenced the genomes of ten species that produce diverse arrays of secondary metabolites in culture. One of the sequenced isolates was described as a new species, and we mapped secondary metabolites detected in culture to the corresponding biosynthetic gene clusters. The ten sequenced genomes were analyzed together with published *Penicillium* genomes, altogether 24, and we developed a pipeline to group biosynthetic gene clusters and map them to known pathways. We found a large untapped potential for biosynthesis of secondary metabolites, encoded in the genomes of these species, that potentially could fill the drug discovery pipeline. Based on our predictions, we experimentally identified a novel compound from the antifungal class of antibiotics called yanuthones.

Since heterologous expression of secondary metabolite pathways has proved troublesome, the ten genome-sequenced *Penicillium* species were evaluated as cell factories in controlled bioreactor fermentations. Compared to an industrially relevant strain, the ten *Penicillium* species showed growth characteristics that encourage further exploration of their industrial potential. Transcriptome analysis of six of the species enabled the identification of a metabolic network that is responsible for precursor formation of secondary metabolites. This network provides important insight into the further industrial development of *Penicillium* cell factories, and could be used in designing metabolic engineering strategies for optimization of secondary metabolite production.

Altogether this thesis provides novel insights into genetic and metabolic aspects of fungal secondary metabolism. Our findings propose that industrial production of secondary metabolites can be effectively established on the basis of native producers. *Penicillium* species constitute a rich source of drug leads, and possess promising physiological characteristics to be established as industrial production platforms.

Keywords: filamentous fungi, *Penicillium*, secondary metabolism, secondary metabolites, antibiotics, systems biology, genomics, next-generation sequencing, cell factories

List of publications

The thesis will be based on the following publications:

I. **Nielsen, J. C.** and Nielsen, J. (2017). Development of fungal cell factories for the production of secondary metabolites: linking genomics and metabolism. *Synthetic and Systems Biotechnology*, 2:5-12.

II. Grijseels, S.*, **Nielsen, J. C.***, Randelovic, M., Nielsen, J., Nielsen, K. F., Workman, M. and Frisvad, J. C. (2016). *Penicillium arizonense*, a new, genome sequenced fungal species, reveals a high chemical diversity in secreted metabolites. *Scientific Reports*, 6:35112.

III. **Nielsen, J. C.**, Grijseels, S., Prigent, S., Ji, B., Dainat, J., Nielsen, K. F., Frisvad, J. C., Workman, M. and Nielsen, J. (2017). Global analysis of secondary metabolite gene clusters reveals vast potential of secondary metabolite production in *Penicillium* species. *Nature Microbiology*, 2:17044.

IV. Grijseels, S., **Nielsen, J. C.**, Nielsen, J., Larsen, T. O., Frisvad, J. C., Nielsen, K. F., Frandsen, R. J. N., Workman, M. (2017). Physiological characterization of secondary metabolite producing *Penicillium* cell factories. *Fungal Biology and Biotechnology*, 4:8

V. **Nielsen, J. C.,** Prigent, S., Grijseels, S., Workman, M., Ji, B., Nielsen, J. (2018). Metabolic regulation of filamentous fungi is tailored for production of secondary metabolites. *Submitted manuscript*

VI. Prigent, S., **Nielsen, J. C.**, Frisvad, J. C., and Nielsen, J. (2018). Automatic reconstruction of 24 *Penicillium* genome-scale metabolic models shows diversity in the secondary metabolism. *Submitted manuscript*

Additional publications not included in the thesis:

VII. Büttel, Z., Díaz R., Dirnberger, B., Flak, M., Grijseels, S., Kwon, M. J., **Nielsen, J. C.**, Nygård, Y., Phule, P., Pohl C., Prigent, S. (2015). Unlocking the potential of fungi: the QuantFung project. *Fungal Biology and Biotechnology,* 2:6.

VIII. Grijseels, S., Pohl, C., **Nielsen, J. C.**, Wasil, Z., Nygård, Y., Nielsen J., Frisvad, J. C., Nielsen, K. F., Workman, M., Larsen, T. O., Driessen, A., Frandsen, R. J. N. (2017). Identification of the calbistrin biosynthetic gene cluster in *Penicillium decumbens*. *Manuscript*

*Contributed equally

Contribution summary

I. Performed literature study and wrote manuscript.

II. Conducted bioinformatics analyses and co-wrote manuscript.

III. Designed study, analyzed data and wrote manuscript.

IV. Conducted fermentations, contributed to data analysis and edited the manuscript.

V. Designed study, analyzed data and wrote manuscript.

VI. Contributed to study design, contributed to network analysis and contributed to writing and editing of the manuscript.

Preface

This dissertation is submitted for the partial fulfilment of the degree of doctor of philosophy. It is based on work carried out between September 2014 and December 2017 in the Systems and Synthetic Biology group, Department of Biology and Biological Engineering, Chalmers University of Technology, under the supervision of Professor Jens Nielsen. The research was funded by the Marie Curie Initial Training Network, QuantFung (FP7-People-2013-ITN, grant no. 607332).

Jens Christian Nielsen

January 2018

Contents

Abbreviations

2MCC	2-methylcitrate cycle
6-MSA	6-methylsalicylic acid
6-MSAS	6-methylsalicylic acid synthase
ARTS	Antibiotic resistant target seeker
BCAA	Branched chain amino acid
BGC	Biosynthetic gene cluster
CAZy	Carbohydrate active enzyme
C	Condensation
COG	Clusters of orthologous groups
CM	Complex medium
DEG	Differentially expressed genes
DM	Defined medium
FBA	Flux balance analysis
GEM	Genome-scale metabolic model
GCF	Gene cluster family
GH	Glycoside hydrolase
GM	Genetically modified
GO	Gene ontology
HGT	Horizontal gene transfer
ITS	Internal transcribed spacer
KOG	Eukaryotic orthologous groups
KS	Ketoacyl synthase
LC-MS	Liquid chromatography-mass spectrometry
NRP	Non-ribosomal peptide
NRPS	Non-ribosomal peptide synthetase
PCA	Principal component analysis
PCC	Pearson correlation coefficient
PK	Polyketide
PKS	Polyketide synthase
NGS	Next-generation sequencing

"Science, my boy, is made up of mistakes,
but they are mistakes which it is useful to make,
because they lead little by little to the truth."

- Jules Verne, A Journey to the Centre of the Earth

Introduction

The discovery of antibiotics is largely associated with the famous and serendipitous discovery of the antibacterial property of a *Penicillium* mold by Alexander Fleming (Fleming 1929). Earlier on, Paul Erlich, had already hypothesized the idea of developing a "magic bullet" that selectively would target disease causing microbes and not the host, based on his observations that certain dyes could selectively stain different cell types (a precursor to gram-staining of bacteria). By applying a systematic screening approach testing the efficacy of synthetic arsenic derivatives on rabbits, Erlich developed Salvarsan in 1909, the first successful treatment for syphilis. Following Flemings discovery, of what today is known as the antibiotic penicillin, a race began on isolating and scaling up the production of this new antimicrobial compound. It was Chain and Florey who published the first purification of penicillin (Chain et al. 1940), and their work led to industrial mass production in 1945, the same year as they together with Fleming shared the Nobel Prize in Physiology or Medicine. As an immediate consequence of the realization that Erlich's "magic bullets" could be derived from microorganisms, researchers started to do systematic screening for antimicrobials from microorganisms in what is known as *the golden era* of antibiotic discovery. One of the pioneers was Nobel Laureate Selman Waksman, who the conducted systematic screening of soil bacteria and identified more than 20 new antibiotics, including some that are in therapeutic use today (Kresge et al. 2004).

Already at an early stage, the problem that bacteria could develop resistance to antibiotics was realized, in particular from synthetic sulfonamide drugs developed in the 1930s. In his Nobel speech, Fleming warned about using too low doses of antibiotics to prevent development of resistance. Nevertheless, the occurrence of antibiotic resistant infectious microorganisms would develop in the years to come for virtually all antibiotics. At the same time, discovery of novel antibiotics started to become increasingly difficult, since most apparently new substances, turned out to be re-discoveries of known antibiotics, thus increasing the costs of development. As a consequence of increased developmental costs, and increasingly strict documentation requirements from the Food and Drug Administration (FDA), many pharmaceutical companies discontinued their development of antibiotics in the 80s and 90s (Davies and Davies 2010). Since the end of the *golden age* in the 1970s, no new classes of antibiotics have been developed (Aminov 2010).

Today, the World Health Organization lists antibiotic resistant pathogens as one of the biggest threats to global health, food security and development (WHO 2017). One important reason for this is the continuous overuse and misuse of antibiotics from healthcare personnel and the agricultural sector. However, in order to reduce the usage of, in particular certain broad-spectrum antibiotics, which will render them useless in the future, there is a need to increase the antibiotics portfolio. This would allow for alternating between different antibiotics and thus distribute the use on a larger number of compounds, in order to reduce the selection for resistance against few highly used antibiotics. Further, more antibiotics would allow for restricting certain classes to human use only, while reserving others for animal treatment, to prevent untreatable pathogens being transmitted between animals and humans.

As a fellow in the Marie Curie ITN, QuantFung, I took part in an effort to unlock the potential of fungi as sources of bioactive compounds (Büttel et al. 2015). In particular, it intrigued me why the development of antibiotics has plateaued and what we can do to fill the antibiotics discovery pipeline. I was curious to investigate if natures reservoir of antimicrobials was exhausted or if unexploited niches could be identified. To study this, we employed genome sequencing of selected fungal species to assess known and unknown biosynthetic pathways, and evaluate the relevance of these pathways for production of novel antibiotics. To facilitate further development of industrial antibiotic production processes, we evaluated the performance of native fungal antibiotics producers as cell factories. This thesis aims at contributing to the development of novel antibiotics from biodiversity screening to industrial exploitation.

Background

Industrial biotechnology

Biotechnology is a broad and multidisciplinary field that revolves around the technological exploitation of biological systems. In a broad sense, this can be expanded to include ancient microbial process such as production of fermented beverages, which dates back to the 7th millennium BC (McGovern et al. 2004). Further it could be argued that low level technological improvement of nature such as grafting of plants can be considered biotechnology. For a more tangible definition, the term biotechnology is commonly divided into different sub-fields defined by colors. The exact definition and number of colors of these sub-fields of biotechnology, are however, not fully agreed upon, and in an editorial in *the electronic journal of Biotechnology*, ten different colors, or sub-fields of biotechnology were defined (DaSilva and J. 2012). More commonly, four colors are used to describe the main fields of biotechnology and each of these fields have some degree of overlap. The four colors of biotechnology described here are green, blue, red and white (**Figure 1A**).

Green biotechnology has to do with applications related to agriculture, e.g. genetically modified (GM) crops. Although controversial, GM plants have the potential to increase nutritional values, sustainability of production and increase resistance to pests (Lucht 2015). Blue biotechnology revolves around exploitation of marine resources, and possess a large potential, since the sea accounts for the greatest biodiversity and area of the earth (Querellou et al. 2010). Red biotechnology constitutes health related applications of biological systems, such as production of pharmaceuticals or biomedical solutions. Lastly, white biotechnology is the fields often referred to as industrial biotechnology and revolves around the production of chemicals using microorganisms or their derived products.

The purpose of industrial biotechnology is to produce chemicals that are easily degradable, requires less energy to produce or plainly perform better than alternatives. These chemicals are produced through biological catalysis using either living cells in a fermentation processes, often referred to as cell factories, or using biologically derived constituents such as enzymes. White biotechnology is widely regarded to represent the next evolutionary step in production processes, and constitutes a sustainable alternative to classical petrochemical production, which is based on microbial conversion of renewable biomass into refined chemical products. The concept of industrial biotechnology is not new and has been applied for decades in the production of selected products such as amylases, citric acid and penicillin, because it has offered a feasible production process of the given products.

Penicillin production is a prime example of biotechnological exploitation in the interface between red and white biotechnology. Penicillin production is carried out by filamentous fungi of the *Penicillium* genus, and has been the focus of great development and optimization through programs of random mutagenesis and selection for beneficial traits.

In recent years however, the technological advancements within several fields of biology, including sequencing and other omics technologies, and gene engineering, have dramatically increased our understanding of cell physiology and genetics. This increased understanding of biological systems can be exploited within the field of metabolic engineering, where rational genetic changes are implemented in cells to improve production capabilities, rather than relying on the stochastic nature of traditional mutagenesis methods. This approach holds a promising potential to change our society to a bio based economy where any chemical in the future could be produced in a feasible way from renewable resources using microbial cell factories.

The field of metabolic engineering aims at applying rational genetic changes to an organism, in order to improve key parameters for industrial production such as the titre, yield and rate (Nielsen and Keasling 2016). Improving these parameters is often undertaken through the iterative cycle of metabolic engineering, where the phenotype of either a native producer or a platform cell factory, expressing a heterologous pathway of interest, is being optimized through the design-build-test-learn cycle (**Figure 1B**). Firstly, a design strategy is developed, which aims at rationally identifying changes of the genetic basis of a production organism that could lead to an improved phenotype. The aim of the design strategy could be to transfer a pathway to a new host, to improve product formation, or to modify properties of the cell such as increasing substrate range (Nielsen 2001). Secondly, the build-step implements the genetic modifications and can benefit from modular synthetic biology parts and the versatility of CRISPR based methods for gene editing (Jakociunas et al. 2016). The physiology of the constructed organism is then evaluated in the test-step e.g. through cultivation experiments and analysis of omics data. In this step, it is tested whether the genetic alterations resulted in the desired traits. Lastly, the learn-step gathers the information from the test-step, in order to gain new insights into physiology and genetics of the organism, and this obtained knowledge can then be exploited in another round of the metabolic engineering cycle by designing a new optimization strategy.

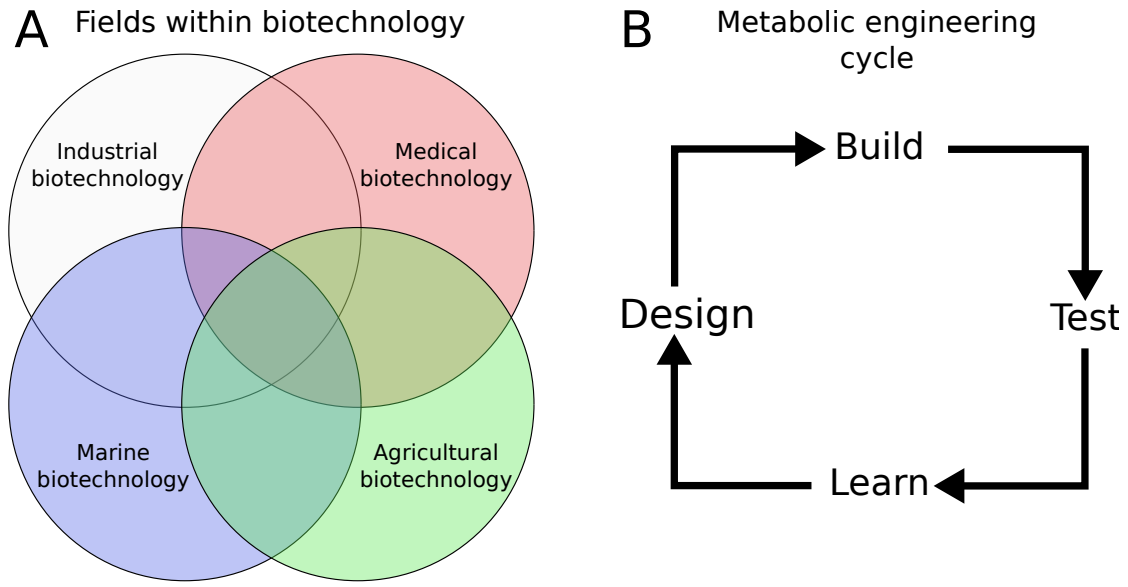


Figure 1. Overview of biotechnology and metabolic engineering. (A) Biotechnology is commonly divided into several partly overlapping sub-fields, which are defined by different colors. (B) The design-build-test-learn cycle of metabolic engineering for optimization of microbial cell factories.

Filamentous fungi

Life-style, growth and taxonomy

Filamentous fungi, or plainly mold in common English, constitute a large and diverse group of organisms within the kingdom of fungi that are characterized by filamentous growth. These organisms are of medical, ecological and industrial importance. In nature, filamentous fungi are ubiquitous, and can be isolated from a wide range of diverse environments and under highly varying conditions in terms of temperature, pH, water activity, etc. Many filamentous fungi are saprophytes, meaning that they live on dead organic matter, which they degrade through the secretion of enzymes, why they also play important roles in the ecosystem as principal decomposers. Another important characteristic affecting the ecosystem is the secretion of bioactive compounds, which are used as a means of communication with other organisms. Colonization of nutrient sources is achieved by growing in a network of branching tubes, hyphae, where transportation of nutrients between cells takes place through septate perforations (Webster and Weber 2007).

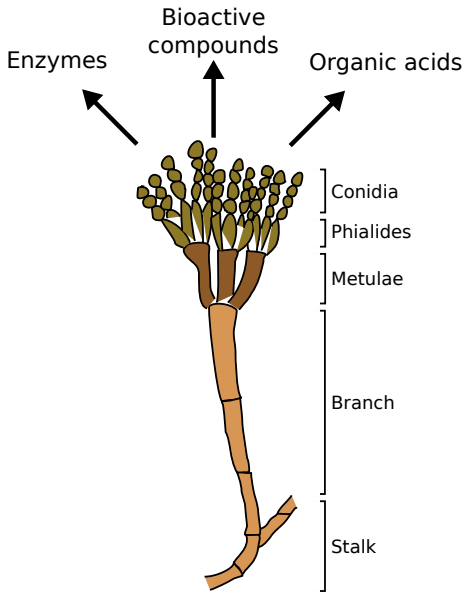


Figure 2. Filamentous fungi represented by a typical conidiophore of a monoverticillate *Penicillium* species. The terminology of different cell types of the conidiophore is indicated, and industrially important products are shown with arrows.

Classification of fungi has been subject to much debate in recent years with sequencing technologies providing new insight into the time of divergence between organisms. Historically, however, fungi are classified based on spore formation characteristics, and they are divided into three classes: zygomycetes, basidiomycetes and ascomycetes. Many of the industrially relevant fungi belong to the group of ascomycetes, including *Saccharomyces* yeasts, and *Penicillium* and *Aspergillus* filamentous fungi. Ascomycetes reproduce either by sexual or asexual reproduction: during sexual reproduction, cells of opposite mating type, fuse and give rise to ascospores. In asexual reproduction, asexual spores called conidia, are generated in sporefoming structures called conidiophores

(**Figure 2**). For many fungi, sexual formation has not been observed, and fungal genera are therefore often defined based on characteristics of the conidiophores. For the genus of *Penicillium*, the brush-like shape of the spore forming units called phialides and conidia is a defining character that also gave rise to the name *Penicillium* meaning "painters brush" (Crous et al. 2009). Based on the number of branch points in the conidiophores, the *Penicillium* genus can be divided into two sub-genera: *Aspergilloides* with monoverticillate conidiophores (no branches) and *Penicillium* with biverticillate or terverticillate conidiophores (one or more staged branch points), and these sub-genera can be further divided into 25 different sections. The sub-genus *Penicillium* is the largest and include several industrially exploited species (Visagie et al. 2014).

Industrial processes

The industrial impact of filamentous fungi is of significant importance especially in the food industry where their usage goes several millennia back (Demain and Martens 2017). Arguably, the most studied genus of filamentous fungi is *Aspergillus*, likely attributable to a long history of industrial exploitation of the koji molds (*A. oryzae* and *A. soyae*) for production of Asian fermented food products. Production of the enzyme, α-amylase in *A. oryzae* was the first patented microbial enzyme (US Patent 525,823, 1894), and industrial citric acid production in *A. niger* dates back a century (Currie 1917). In addition, *A. fumigatus*, is a human pathogen causing Aspergillosis in immunocompromised individuals (Latgé 1999), and *A. nidulans* has been widely used as a model system to study filamentous fungi since the 1950s (Pontecorvo et al. 1953).

The *Penicillium* genus is phylogenetically closely related to Aspergilli and is also being widely used in industrial settings. Industrial applications of Penicillia include: (*i*) manufacturing of fermented food products such as blue veined cheeses using *P. roqueforti* and bloomy rind cheeses using *P. camemberti* (Cheeseman et al. 2014), (*ii*) *P. nalgiovense* is used as a starter culture in fermented dry sausages (Ludemann et al. 2009), (*iii*) production of lignocellulolytic enzymes (Liu et al. 2013a) and (*iv*) production of bioactive small molecules, which can be used in the pharmaceutical industry (Frisvad et al. 2004). Further, Penicillia play an important role in mediating phosphate to plants (Chai et al. 2011), which has great impact on agriculture (Richardson and Simpson 2011). As previously mentioned in the Introduction, it was a *Penicillium* colony that Fleming found as a contaminant on his bacterial agar plates, and the genus have since attracted attention owing to its potential for production of drug leads from the group of compounds referred to as secondary metabolites.

Secondary metabolism

Introduction to secondary metabolism

Metabolism can be broadly defined as the set chemical transformations taking place inside living cells, with the majority of these being catalyzed by enzymes (Nielsen 2017). Arguably, the overshadowing goal of metabolism is to generate the necessary building blocks that allow an organism to proliferate. This is achieved through uptake of nutrients and transformation of these into energy and precursor units for the synthesis of

macromolecules such as proteins, nucleic acids, lipids and carbohydrates. These essential processes comprise, in most organisms, what is usually referred to as metabolism.

However, some species, often constrained to specific taxonomical lineages, have evolved additional metabolic features, which are not essential for growth or survival in the organism, and hence is referred to as secondary metabolism (this definition has proved limited, although it mostly holds true). The chemicals resulting from the secondary metabolism are referred to as secondary metabolites or natural products, and constitute a group of structurally diverse compounds. Some of the most prolific secondary metabolite producers include *Actinobacteria*, plants and filamentous fungi (Bérdy 2005). The ecological function of most secondary metabolites is unknown, but they are generally believed to be used as a means of communication, and to confer the producing organism with a fitness advantage that renders it better adapted to a given, often stressful, environment (Romero et al. 2011). Sometimes the ecological function of secondary metabolites can be deduced from the activity of the compounds. For example, some secondary metabolites are pigments, like melanin in *Penicillium* species, which is responsible for the coloring of the spores. This has been hypothesized to be a means of blocking ultra-violet irradiation from the sun, to avoid mutations (Butler and Day 1998). Other secondary metabolites such as terrein from *A. terreus* are chelators, and function as metal scavengers that are produced in nutrient scarce environments (Gressler et al. 2015). Lastly, many secondary metabolites exhibit bioactivities and are likely a means to fight off predators or competitors by killing them or inhibiting their growth (Rohlfs and Churchill 2011).

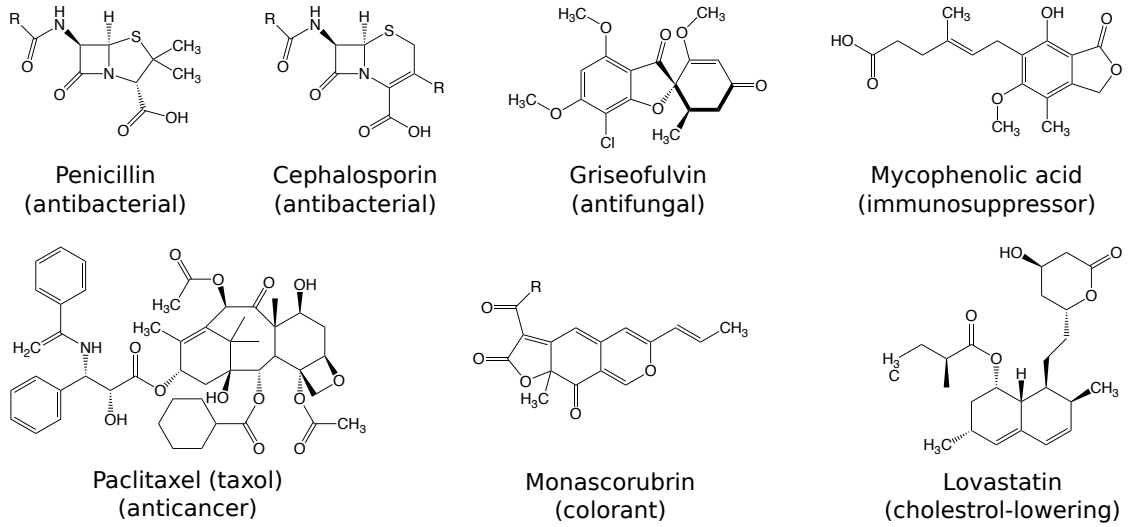


Figure 3. Industrially important fungal secondary metabolites. Aside from antibiotics, fungal secondary metabolites are used as other pharmaceuticals such as immunosuppressants and anticancer drugs, and as food colorants.

Industrial applications of secondary metabolites

From a pharmaceutical point of view, secondary metabolites with biological activity are interesting candidates for drug leads (**Figure 3**). Many compounds from fungi possess antimicrobial activities such as the β-lactam antibiotics penicillins and cephalosporins (Brakhage 1998), as well as the antifungal griseofulvin (Finkelstein et al. 1996).

Pharmaceutical applications of fungal secondary metabolites, however, goes beyond antibiotics and include immunosuppressants such as mycophenolic acid (Stassen et al. 2007) and cyclosporins (Sallam et al. 2005), cholesterol lowering drugs such as statins (Barrios-González and Miranda 2010) and anticancer drugs such as taxol (Yang et al. 2014). Aside from pharmaceutical properties, some fungal secondary metabolites can be exploited as natural colorants, such as monascorubrin and its derivatives which are produced by *Monascus* species and are widely used as dyes in the food industry (Woo et al. 2014). Altogether, fungal derived secondary metabolites constitute an important multi-billion dollar industry (Hillman et al. 2017).

Secondary metabolite biosynthesis

One of the explanations for the diverse industrial applications of secondary metabolites, relates to the structural diversity of the chemicals (**Figure 3**). In spite of this diversity, secondary metabolites are derived from a limited number of precursors from the primary metabolism. The main fungal secondary metabolites are polyketides (PK) derived from short chain carboxylic acids, and non-ribosomal peptides (NRPs) derived from amino acids (Keller et al. 2005; Nielsen et al. 2017) (**Figure 4**). Secondary metabolites are polymers of these precursors, that are linked together using polyketide synthases (PKSs) or non-ribosomal peptide synthetases (NRPSs) and the resulting chemical structures are subsequently modified by various tailoring enzymes such as oxygenases and transferases, to confer additional structural functionality to the final product. Secondary metabolite pathways are often coexpressed with a transporter to facilitate secretion of the final product and to detoxify the cell from bioactive intermediates or end-products. Other self-protection mechanisms include genes encoding an unsusceptible version of the enzyme targeted by the chemical products of the pathway (Regueira et al. 2011). The genes encoding secondary metabolite biosynthesis pathways tend to cluster in the genome in biosynthetic gene clusters (BGCs) (**Figure 5A**).

The precursors for secondary metabolites are highly connected in metabolism where they are used for the biosynthesis of macromolecules, energy and co-factors (Nielsen 2014). Specific environmental stimuli can activate expression of specific secondary metabolite pathways, and direct carbon flux towards their biosynthesis (**Figure 4**). This is controlled by transcription factors, such as the global transcriptional regulator, the velvet complex (VeA), that regulates both secondary metabolism as well as other processes such as sexual development (Bayram et al. 2008). Such global regulators further control the expression of either specific pathways or other transcription factors that are pathway specific and often belong to the zinc cluster family of transcription factors (Shelest 2008). These pathway specific regulators also ensure concerted expression of pathway genes. In addition to PKS and NRPS pathways, less frequent, but equally important secondary metabolite pathways include the ones synthesizing terpenoids, alkaloids, as well as hybrids between different classes (Hoffmeister and Keller 2007).

There are a number of metabolic routes generating acetyl-CoA and amino acids in fungi, and since many secondary metabolites are only produced during starvation; breakdown might be an important mechanism to generate precursors for the biosynthesis of secondary metabolites (**Figure 4**). Roze et al. (2010) suggested the breakdown of branched chains amino acids (BCAAs) (isoleucine, leucine and valine) and fatty acids, through β-oxidation to be some of the main contributing pathways that provide acetyl-CoA for the biosynthesis of PKs in Aspergilli (Roze et al. 2010).

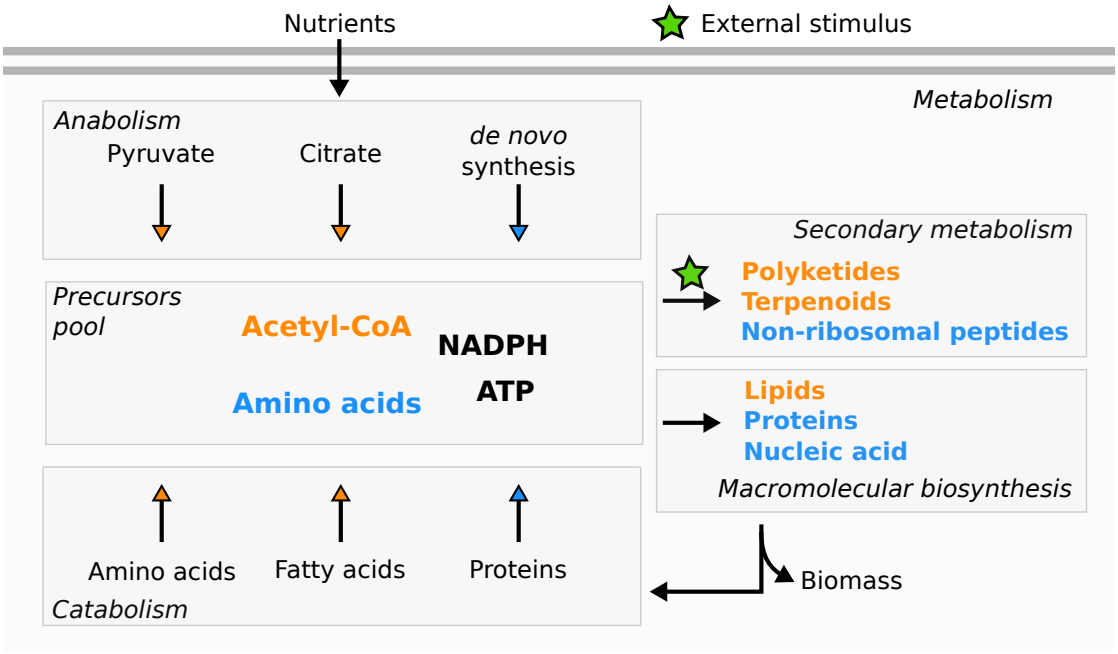


Figure 4. Overview of metabolism. Nutrients are taken up and converted into precursor metabolites through the process of anabolism. These precursors are transformed into macromolecules to generate biomass for growth. During nutrient limited conditions precursors can be regenerated through catabolism of macromolecules and repurposed in the cell. External stimuli can induce secondary metabolism and direct flux of precursor metabolites towards generation of secondary metabolites. Acetyl-CoA and amino acids are important building blocks for macromolecules and secondary metabolites.

Architecture of biosynthetic enzymes

The PKS and NRPS backbone genes, which products catalyze the committed step in secondary metabolite biosynthetic pathways, are modular and multidomain megaenzymes. PKSs are divided into different sub-classes depending on the specific mechanism of action. Fungal PKSs are type 1 iterative PKSs, which consist of a single module of catalytic protein domains that work in an iterative fashion and are evolutionarily related to vertebrate fatty acid synthases (Smith and Tsai 2007). In a minimal type I PKS, the protein domains include an acyltransferase (AT), a ketoacyl synthase (KS), an acyl-carrier protein (ACP), a starter ACP transacylase (SAT) and a thioesterase (TE) (**Figure 5B**). The starter unit, often acetyl delivered by the enzyme CoA, is recognised by the SAT domain and then loaded onto the KS domain. An extender unit, often malonyl delivered in its CoA form, is bound by the AT domain and fused to the starter unit through a C-C bond forming Claisen condensation catalyzed by the KS domain and driven by the decarboxylation of the extender unit. This acyl chain is then transferred to ACP that moves the growing carbon backbone between active sites. Either another round of elongation is performed, or the acyl chain is transferred to the TE domain and released from the enzyme. The length of the growing chain, and thus the number of iterations performed has proved to be correlated to the volume of the active site of the KS domain (Yadav et al. 2009). The resulting non-reducing PKs forms aromatic structures.

The optional part of PKSs include the reducing domains enoyl reductase (ER), dehydratase (DH) and β-ketoacyl reductase (KR). These domains act on the β-keto group of the acyl chain, which can be partially reduced or fully reduced, based on the presence of these domains in the PKS and the activity of the domains during different iterations.

Thus, during chain elongation, the keto group on the β-carbon is either retained, reduced to a hydroxyl group or reduced to an enoyl group along the acyl chain (**Figure 5B**). The main difference between PKSs and fatty acid synthases is that reduction of the β-carbon is optional on PKSs.

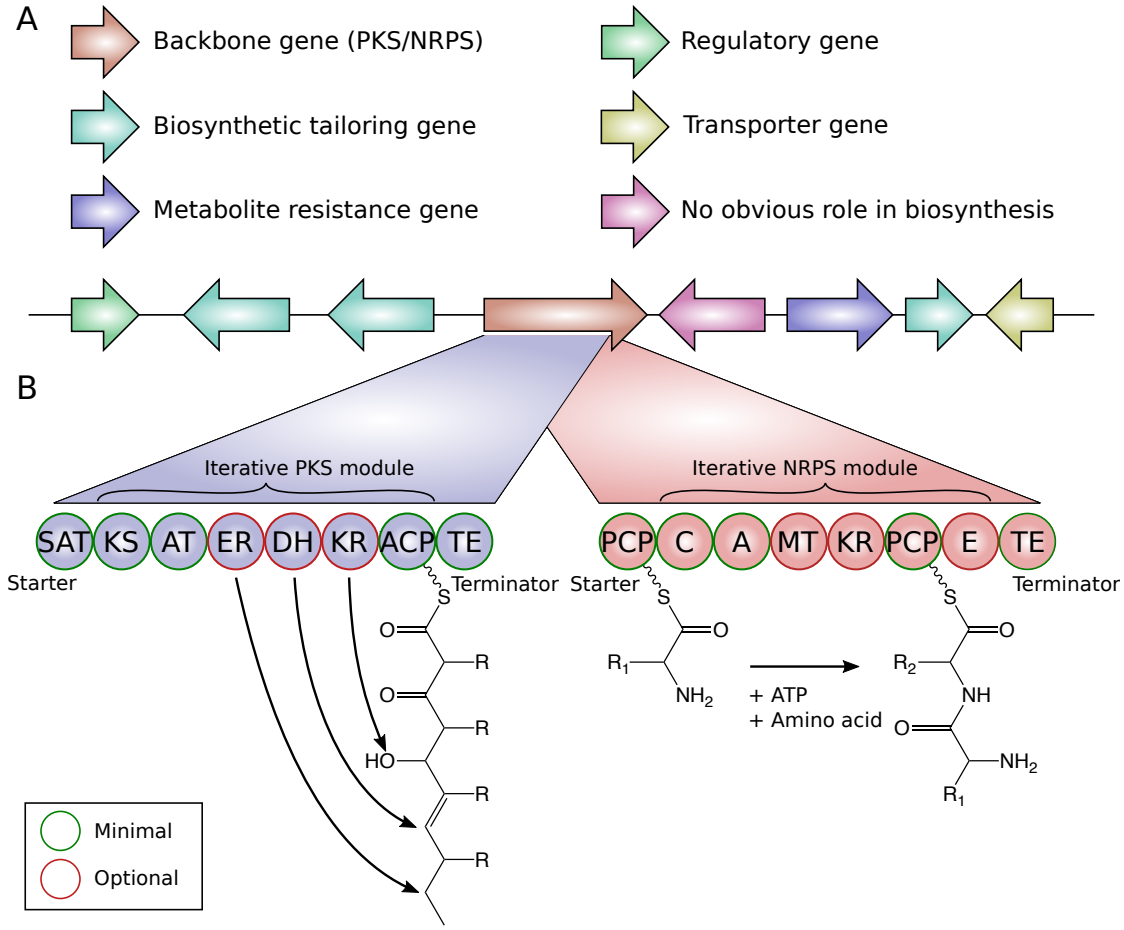


Figure 5. Overview of secondary metabolite biosynthetic gene clusters (BGCs). (A) Typical BGC and the most common classes of genes present. (B) Protein domain architecture and catalytic principle of PKS and NRPS backbone genes catalyzing the formations of PKs, and NRPs, respectively. See text for abbreviations of protein domains.

NRPSs are characterized by their large size, which in most organisms constitutes the longest amino acid sequences in the proteome, being several thousand amino acids long. Their modular organization show similarity to that of PKSs, although substrates differ considerably in being amino acids. A minimal NRPS contain a condensation (C) domain, an adenylation (A) domain, a peptidyl carrier protein (PCP), and a 4´phosphopantetheine (4´PP) dependent transferase (PPT) (**Figure 5B**). The A domain is amino acid specific and selects a given amino acid, that is activated by hydrolysis of ATP to form an aminoacyl adenylate, which is then bound to a 4'-PP cofactor thiol group attached to the PCP domain. Two aminoacyl adenylates are fused by the C-C bond forming condensation reaction catalyzed by the C domain (Sieber and Marahiel 2005). NRPS modules can work in an iterative manner where the same A, C and PCP domains continuously increase the peptide chain length, or a number of modules can work in sequence to yield the final peptide (Yu et al. 2017). Optional domains include methyltransferases (MT), and

epimerases (E) that are responsible for methylation and chirality changes of the amino acids, respectively.

Gene clustering

As previously mentioned, a central trait for the genes encoding secondary metabolite biosynthetic pathways, is that they tend to physically cluster in a single locus in the genome in BGCs (Smith et al. 1990). Gene clustering of metabolic pathways is not common in eukaryotes, possibly because of the associated risk of losing an entire essential pathway if part of a chromosome is lost. However, in bacteria, pathway genes are often clustered in operons, which has been proposed to be attributed to a selfish property of the genes, as they are more likely to survive horizontal transfer if all pathway members are acquired at once (Lawrence and Roth 1996). It has been shown that many fungal BGCs originate from bacteria and have been acquired through horizontal gene transfer (HGT), and thus could explain the gene clustering in fungi (Wisecaver and Rokas 2015). Since secondary metabolites are not essential, there might not be a strong selective pressure to distribute the pathway genes in the genome. The gene clustering might also serve as an advantageous trait for rapidly evolving new secondary metabolite pathways. The genomic loci of BGCs have been proposed as "evolutionary playing fields" where mutations would allow rapid evolution to generate new and useful compounds, without the risk of harming the host by disrupting essential genes (Lind et al. 2017). In support of this notion, it has been shown that BGCs often localize in the sub-telomeric region of the chromosomes (McDonagh et al. 2008; Palmer and Keller 2010), which are regions with a low frequency of essential genes, hence allowing for evolutionary changes to occur without harming the host.

Genome mining of secondary metabolism

The genome era reached filamentous fungi in the mid 2000s, with *Neurosporra crassa* being the first genome sequenced representative in 2003 (Galagan et al. 2003). This was followed by the release of three *Aspergillus* genomes in a single issue of *Nature* in 2005 (Galagan et al. 2005; Machida et al. 2005; Nierman et al. 2005). An important finding emerged from the analysis of these genomes, as it was clear that the number of secondary metabolite BGCs was far greater than the number of compounds produced, effectively demonstrating a yet unexploited biosynthetic potential encoded in the genomes of these organisms. This led to the notion that the majority of secondary metabolite BGCs are silent under standard laboratory conditions (Bok et al. 2006).

Following the discovery of silent BGCs, a number of different bioinformatics tools have been developed to automatically identify BGCs in genomes (Medema and Fischbach 2015; Ziemert et al. 2016). The bioinformatics approaches for the discovery of BGCs, generally, utilize the convenient trait of gene clustering to identify all genes encoding a single pathway. By sequence analysis, e.g. using Hidden Markov Models (HMMs), a genome can be screened for known backbone genes such as PKSs, NRPSs, terpene cyclases etc. From these, the extent of a BGC can be estimated based on the presence of common biosynthetic genes in the vicinity of the backbone gene. These algorithms, identifies, with high confidence and fully automated, BGCs of the common classes, PKSs, NRPSs etc., although prediction of BGC boundaries is troublesome and determination of the gene members encoding a pathway requires manual curation and/or

experimental investigations. A number of tools have been implemented using the strategy described above with SMURF (Khaldi et al. 2010) and antiSMASH (Medema et al. 2011) as the most successful examples, and recently the cluster boundary predictions have been improved by incorporating analysis of conserved promoter motifs in fungal BGCs (Wolf et al. 2016). Other types of algorithms for BGC prediction are motif independent and thus allow for the identification of unknown classes of BGCs, either by investigating coexpression patterns of clustered genes (Andersen et al. 2013; Umemura et al. 2013), comparative genome analysis (Takeda et al. 2014) or machine learning strategies based on the frequency of PFAM domains (Cimermancic et al. 2014). With genome mining for BGC identification, secondary metabolite research has opened up for applying systems biology tools in the study of secondary metabolism.

Systems biology and omics

The paradigm of systems biology

Systems biology is a computational discipline that aims at studying biological systems as a whole and applies data integration and mathematical modeling to gain insight (Nielsen 2017). Being it interactions between organisms in a biological niche, interactions between tissues in higher organisms or interactions of enzymes and metabolites in a cell, systems biology integrates global information to identify emerging properties of the system. The concept of emerging properties is a central dogma of systems biology, and defines the phenomena occurring from interactions of the components of a biological system that cannot be deduced based on the function of the components alone. This is important because biological systems rarely consist of individual parts that act autonomously, but rather is composed of a range of heterogenic constituents that are interdependent and interconnected. The systems biology paradigm, as defined by Kitano (2002a), represents a holistic approach and stands in contrast to traditional reductionist biology, that dissects biological systems and studies their individual components (Kitano 2002a). Reductionism is a useful approach to describe the constituents of a system, but is limited in explaining the system as a whole. In this thesis, the biological system in focus is the cell. Reductionist biology has, through the use of molecular biology, provided insights into the function of cellular constituents such as genes, transcripts, proteins, metabolites, compartments and membranes, and this has provided the foundation for systems biology studies that integrates the accumulated knowledge to study the cell.

Systems biology is often divided into two different approaches: the bottom-up and the top-down approach. The bottom-up approach derives detailed models, e.g. about a biochemical pathway, and requires manual curation and a thorough prior description of the individual components of the system. The top-down approach has been used in this thesis and integrates systems level characterization or quantifications of collections of biological entities, from high throughput technologies, what is often referred to as omics data. Omics data can represent a multitude of different biological sources and techniques, and one of the most established omics technologies is transcriptomics, which quantifies all mRNA transcripts in a single cell or a population of cells, the transcriptome. Similarly, other types of omics technologies include genomics, proteomics and metabolomics. The inclusive and integrative nature of top-down systems biology and the identification of emerging properties, makes it a data-driven discipline that is hypothesis generating. A

common example would be the comparison of cellular responses to different conditions and applying statistical and clustering methods to extrapolate patterns in the data that define a given perturbed state of the cell. The hypothesis generated from systems biology needs to be validated experimentally and the experimental results can then be used to improve the description of a biological system that has increased predictive or descriptive power, and can be applied for further systems level studies. Kitano (2002b), described this as the systems biology research cycle (Kitano 2002b).

Integrative analysis of omics data

As a consequence of technological developments, the generation of omics data is getting cheaper and faster. In particular within nucleic acid sequencing where the massive parallelisation of small sequencing tasks represents what is referred to as next-generation sequencing (NGS). The sequencing of short DNA fragments, provides information about parts of a nucleotide sequence, and the connections between these fragments can be inferred using computational algorithms, that use nucleotide overlaps to assemble the original sequence.

These NGS techniques allow for generation of a range of omics data, such as metagenomics, epigenomics and exome sequencing, which are all different variations of applying NGS to different samples or enrichments of samples (Rizzo and Buck 2012). Most commonly, however, NGS is used for genome sequencing and transcriptome sequencing (RNA-seq).

Genome sequencing provides information on the order of nucleotides in a genome, a chromosome or part of a chromosome referred to as a scaffold or contig. Often, the aim is to identify coding regions of a DNA sequence in order to functionally annotate the genes based on homology to other sequences. A useful application of a genome is to benchmark it against other genomes, what is known as comparative genomics. Here the presence, absence or variation in sequences can form the basis for identifying function or phylogeny between genes or genomes. Comparative genomics and phylogenomics can be used in identifying pathways by mapping phenotypic data to phylogenetic trees, and correlating this with gene presence or absence. Fungi are particularly useful for this due to their manageable genome size, evolutionary divergence and phenotypic variation (Sardi and Gasch 2017).

RNA-seq is probably the most established omics technology used to study the dynamic aspects of cell physiology. This owes largely to the price of sequencing and the standardization of the associated analysis protocols. Transcriptome sequencing data can either be mapped to a reference genome, and thus benefits from previous annotation efforts, or it can be assembled *de novo*. The most common application of transcriptome data is statistical analyses, such as differential expression analysis that identifies the main affected genes when comparing different conditions. Given the complexity of cellular physiology, isolated functional information about genes differentially expressed, provide limited information of the cell as a system, why contextualization in gene sets is useful in the evaluation of the cellular response to a treatment.

The Gene Ontology (GO) terms, aims at providing a vocabulary for gene products, that is dynamic and generally applicable across organisms, and it is frequently used in analysis of omics data (Ashburner et al. 2000). GO terms are organized as a directed acyclic graph where each term has a defined relationship to connected terms and thus is representing a hierarchy of information. Other annotations include Clusters of Orthologous Groups

(COGs) describing bacterial gene products (Tatusov et al. 2000) or the corresponding eukaryal annotation, euKaryotic Orthologous Groups (KOGs) (Tatusov et al. 2003). A number of databases exist, that link genes to metabolic pathways or reactions, such as the Koyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al. 2016) and the MetaCyc database (Caspi et al. 2014). Such mapping of gene functions can be used in connection with enrichment analysis to assess which cellular functions are driven by a group of genes, e.g. the differentially expressed (Väremo et al. 2013).

Metabolic modeling

Metabolism constitutes a network of interwoven biochemical reactions that are not easily interpreted by humans. Given sufficient prior knowledge of an organism, this network can be reconstructed at the genome-scale, and the metabolic flux through such a network can be simulated with computers in genome-scale metabolic models (GEMs) (Price et al. 2004). The foundation of a GEM is the functional annotation of the genes in an organism, and connecting these to the biochemical reactions catalyzed. This can be arranged in a stoichiometric matrix that provides a comprehensive and quantitative description of the metabolic capabilities of a cell and associates genes to reactions and metabolites. Flux balance analysis (FBA) is a modeling concept that can be used to simulate the fluxes in a metabolic network given an objective function. The objective function can vary, but commonly an artificial biomass function is maximized, based on the idea that microorganisms have evolved to maximize their growth rate (Orth et al. 2010). Aside from simulations, GEMs can be used for topological analysis of metabolism, by integrating omics data. This allows for extracting subnetworks that are metabolic hotspots of the perturbations studied. One example of this, is the extraction of reporter features, such as metabolites, that are overrepresented among the reactions catalyzed by affected genes of interest, e.g. differentially expressed genes (DEGs) (Patil and Nielsen 2005).

Part I: Mapping the biosynthetic potential of *Penicillium* using genomic

Within recent years, the field of fungal genomics has experienced exciting development and an increasing level of attention. A main driver of this has been the 1000 fungal genomes project which was initiated with the aim of exploring the genomic diversity within the fungal kingdom (Grigoriev et al. 2014). As part of the project, sequencing of the *Aspergillus* genus constituting approximately 350 species is being undertaken (Brandl and Andersen 2017). This will provide evidence about the diversity of this industrially and medically important genus. A recent genomic study of 19 different *Aspergillus* species provided a thorough investigation into several aspects of *Aspergillus* biology, and linked phenotypes to genotypes for a number of traits (de Vries et al. 2017). In particular, secondary metabolite BGCs were identified across species, and there was overall a good correlation between production of secondary metabolites and identification of the corresponding BGCs.

In comparison, the *Penicillium* genus is understudied at the genomic level, and this is surprising considering the industrial importance of Penicillia. As part of this PhD project, we performed genome sequencing of ten *Penicillium* species. The species were selected in a collaboration with Professor J. C. Frisvad and Associate Professor M. Workman from the Technical University of Denmark (DTU), with the aim of capturing the full spectrum of phylogenetic diversity of the *Penicillium* genus and to select species with strong capabilities for production of diverse arrays of secondary metabolites.

The genome sequencing provided the basis for investigations into *Penicillium* genomics with particular focus on secondary metabolism. In this part of my thesis, I will describe these efforts which constituted analyses to obtain a holistic view of the biosynthetic potential of Penicillia through genome mining. The contents of three publications which have revolved around this topic, will be described in the following, starting with a review on different strategies for linking BGCs to secondary metabolites (**Paper I**). This is followed by a case story where we combined metabolite profiling and genome sequencing to identify BGCs responsible for the biosynthesis of secondary metabolites produced by a new fungal species, *Penicillium arizonense* (**Paper II**). Lastly, I will describe a genus wide investigation of *Penicillium* where we analyzed the diversity of the secondary metabolism across 24 different *Penicillium* species, and used this to gain insights into the evolution of BGCs and production of novel secondary metabolites (**Paper III**).

Paper I: Linking genes to secondary metabolites

The combination of an increasing number of sequenced genomes, and the development of automated genome mining algorithms for the identification of BGCs, has sparked the community with excitement that a second *golden age* of antibiotics discovery is on the rise (Medema and Fischbach 2015). This promise has yet to be fulfilled, and part of the explanation is the challenge of the down-stream analysis of linking identified BGCs to a metabolic pathway or chemical end-product. In **Paper I**, we reviewed different strategies for linking BGCs to pathways and define them as: targeted or untargeted approaches.

Targeted approaches

When the aim is to connect a single BGC to a compound (or *vice versa*), targeted approaches can be applied. This is low-throughput and benefits from expert knowledge on biosynthetic mechanisms.

One simple approach is to identify two or more genome sequenced organisms producing a compound of interest, and then evaluate whether these organisms share any backbone genes that could be responsible for the synthesis of the given class of compounds i.e. PKSs for PKs, NRPSs for NRPs etc. Shared backbone genes between organisms can be identified by sequence similarities of either the full translated protein sequence or based on specific conserved protein domains (Cacho et al. 2015). Another approach is to use retrosynthesis, which aims at deducing the specific enzymes needed for the synthesis of a given compound, for example by evaluating if a PK is reduced, and comparing that to the number of reducing domains in PKSs (Cacho et al. 2015). A combination of comparative genomics and retrosynthesis is often needed to pinpoint a BGC of interest. These targeted approached were successfully used in the identification of the BGCs responsible for the synthesis of the industrially important secondary metabolites griseofulvin (Chooi et al. 2010) and tryptoquialanine (Gao et al. 2011) in *P. lanosocoeruleum*.

Another piece of information that can be used in deducing a match between a BGC and a compound is the presence of self-resistance genes, which have been proved to co-localize with the pathway genes in some BGCs. This information was utilized to identify a BGC responsible for mycophenolic acid biosynthesis in *P. brevicompactum* where a duplicated version of the mycophenolic acid target enzyme, IMP dehydrogenase, encoded an unsusceptible version that conferred self-resistance to the organism (Regueira et al. 2011). More recently, a BGC in *A. nidulans* containing a gene encoding an extra copy of a proteasome subunit was identified, and the end-product of the pathways proved to be the proteasome inhibitor fellutamide B (Yeh et al. 2016).

Untargeted approaches

Above-mentioned targeted approaches are labour intensive and unfeasible for exhaustive annotation of BGCs from many genomes. Thus, an appealing thought is to generalize the targeted approaches, through algorithmic implementation. Recent studies have made great progress in development of computational retrosynthesis, which has been integrated in an automatic workflow to correlate gene structures to enzymatic activities encoded in

BGCs (Dejong et al. 2016). Similarly, identification of resistance genes has been suggested to be exploitable for genome mining of BGCs encoding compounds with specific properties (Tang et al. 2015; Ziemert et al. 2016). This approach is applied by the Antibiotic Resistant Target Seeker (ARTS) that use self-resistance genes to prioritize computationally detected BGCs with putative antibiotic activity (Alanjary et al. 2017).

Annotating the function of BGCs has been enabled in recent years with the development of well-annotated databases with sequence information of BGCs. The most inclusive database is IMG-ABC from the Joint Genome Institute (JGI), which contains more than one million automatically mined BGCs, however the majority of these are orphan (Hadjithomas et al. 2015). Conversely, the MIBiG database contains only BGCs that have been linked to an end-product, but the total number of BGCs is also considerably smaller (1393 BGC) (Medema et al. 2015). The MIBiG database has further defined the minimum information about a BGC, which ensures appropriate annotation of new BGCs, enables efficient parsing of the database and facilitates user entry of new items, thus reducing the maintenance and increases the chance of continuation and growth of the database. Another community driven database was previously attempted with the clustermine360 database (Conway and Boddy 2013), but maintenance seem to have been discontinued. Reflecting the literature, these databases are dominated by bacterial entries, however, a dedicated effort to include more fungal sequences used text mining to add an additional 197 BGC of fungal origin to the MIBiG database (Li et al. 2016).

In order to utilize such databases to group BGCs, a similarity metric is needed. Early genomics work on the comparison of fungal type 1 PKSs was conducted by Kroken et al. (2003) who showed that the KS domains of PKS were conserved and well suited for inferring the phylogenetic relationship between PKSs (Kroken et al. 2003). Similarly, C domains of NRPSs have shown to be informative with respect to enzyme architecture and function (Rausch et al. 2007). Ziemert et al. (2014) used these KS and C domains to assess the similarity between PKSs and NRPSs in bacteria of the *Salinispora* genus (Ziemert et al. 2014). More advanced methods enables comparison of BGCs beyond the ones containing PKSs and NRPSs, and use the number of shared PFAM domains between BGCs, (Cimermancic et al. 2014), or a combination of different similarity metrics (Doroghazi et al. 2014). This allows for grouping of BGCs into gene cluster families (GCFs) encoding the same or related pathways. Such GCFs can be mapped to database entries to annotate the corresponding pathways.

In **Paper II** we applied a targeted approach to link secondary metabolites detected in culture extracts, to the corresponding BGCs in *P. arizonense*. In **Paper III** we applied an untargeted approach to link BGCs in 24 species, to chemical compounds, using the MIBiG database.

Paper II: The secondary metabolism of *Penicillium arizonense*

Genome sequencing was conducted of a fungal isolate that we proposed as a new species within the *Penicillium* genus: *Penicillium arizonense* Frisvad, Grijseels and J.C. Nielsen, sp. nov. The isolate was obtained from red soil in Grand Canyon, South Rim, Arizona, USA in July 1990. We described the morphology of the species and conducted a genomic and chemical characterization that indicated a large potential of *P. arizonense* for production of new enzymes and secondary metabolites.

A new species within section *Canescentia*

Preliminary morphological analysis suggested that *P. arizonense* was related to species from section *Canescentia*. We decided to assess its phylogeny compared to other species of this section, based on four nucleotide marker sequences, the internal transcribed spacer (ITS) region, the β-tubulin (*BenA*) gene, the calmodulin (*CaM*) gene and the RNA polymerase II second largest subunit (*RPB2*) gene, which are recommended for assessing *Penicillium* phylogeny (Visagie et al. 2014). A concatenated maximum likelihood tree based on the marker sequences confirmed *P. arizonense* as grouping in section *Canescentia* with *P. yarmokense* as closest relative (**Figure 6**).

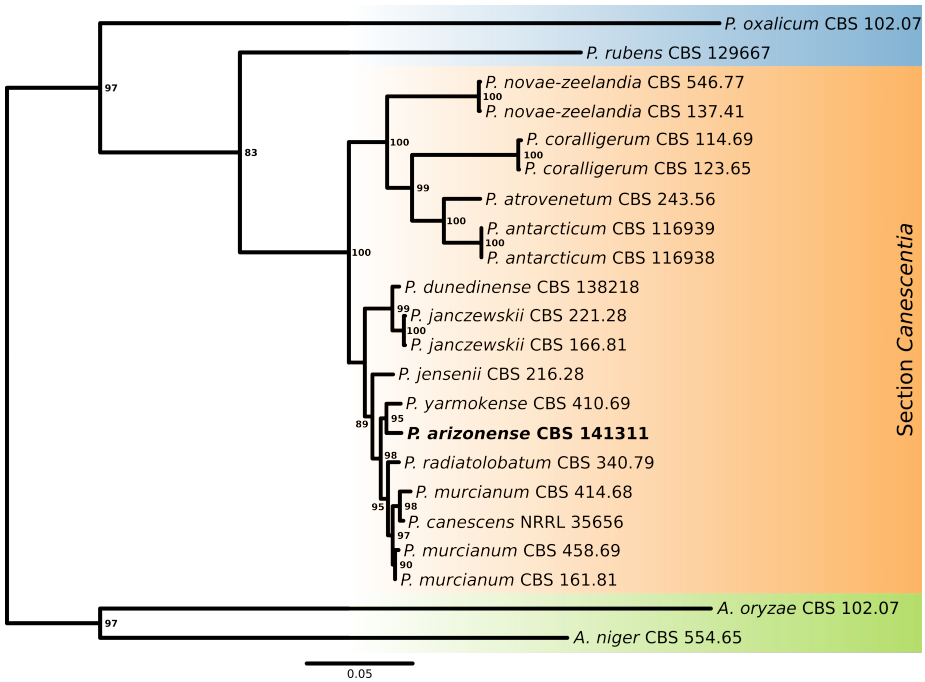


Figure 6. Phylogenetic tree of *Penicillium* species in section *Canescentia*. The phylogeny was inferred as a maximum likelihood tree using the concatenated sequence of four nucleotide marker sequences (ITS, *BenA*, *CaM* and *RPB2*). Bootstrap support is given as percent based on 1000 bootstrap replicates and only indicated in nodes having more than 80% bootstrap support. Scale bar indicates the mean expected substitutions per site.

The *P. arizonense* genome is the first published representative of section *Canescentia*, and the paper is to the best of our knowledge, the first description of a new fungal species

which is published together with its genome sequence. This sets a standard for future species description efforts to provide genome sequencing information as well.

Large arsenal of degradative enzymes

Section *Canescentia* has proved to contain species that are efficient producers of lignocellulolytic enzymes, in particular xylanases as seen in *P. canescens* (Bakri et al. 2003) and *P. janczewskii* (Terrasan et al. 2010). In order to evaluate if *P. arizonense* possessed potential for production of similar industrial enzymes, we annotated Carbohydrate Active enZymes (CAZys) in its genome as well as in related species known for production of hydrolytic enzymes or model organisms (**Figure 7**).

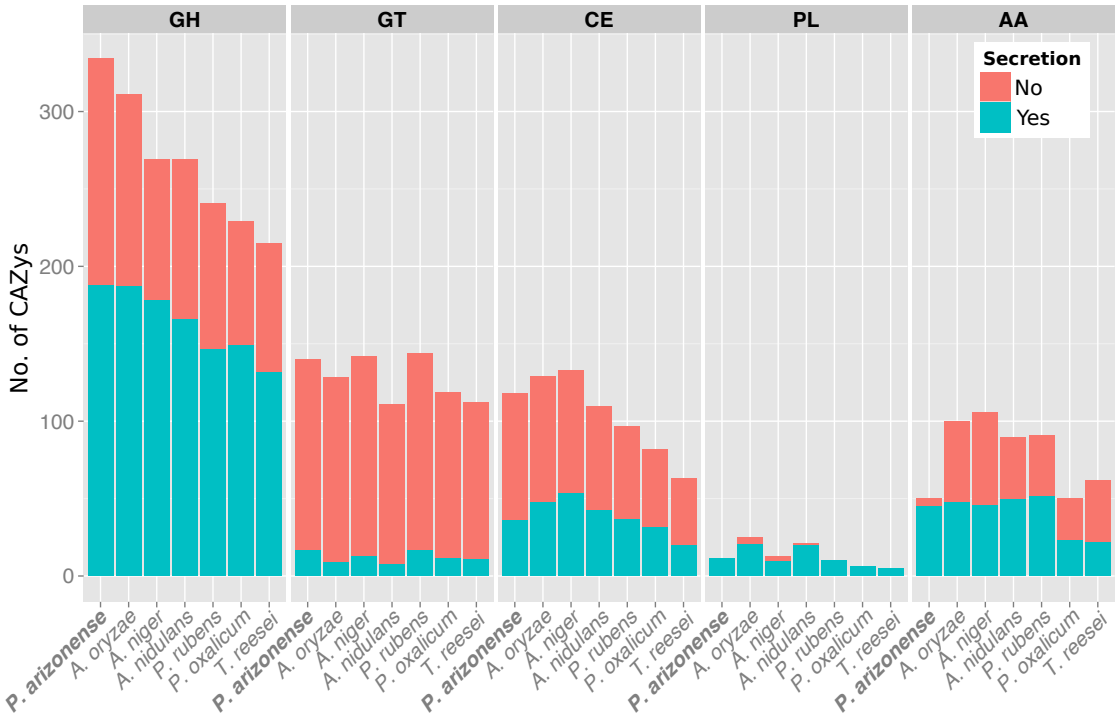


Figure 7. Carbohydrate Active enZymes (CAZys) in filamentous fungi. For all CAZys the presence of a secretion signal was detected, grouping them into secreted or non-secreted. GH: Glycoside Hydrolase; GT: Glycosyl Transferase; CE: Carbohydrate Esterase; PL: Polysaccharide Lyase; AA: Auxiliary Activities.

Interestingly, *P. arizonense* was the species with the highest number of glycoside hydrolase (GH) proteins encoded in its genome. GHs breaks glycosidic bonds between carbohydrates, and are of relevance for degradation of complex biomass e.g. for production of biofuels (Liu et al. 2013a). Based on evaluation of secretion signals, *P. arizonense* was further found to contain the highest number of secreted GHs, together with *A. oryzae* whose capabilities to secrete enzymes is industrially exploited (Christensen et al. 1988). The most important industrial producer of cellulases *Trichoderma reesei*, proved to encode few GHs in agreement with previous observations (Martinez et al. 2008). Recently, Penicillia have been suggested to be promising producers of lignocellulolytic enzymes, due to increased protein production and hydrolytic performance of *Penicillium* enzymes compared to *T. reesei* (Gusakov 2011). Whether the high number of GHs identified in *P. arizonense* have high hydrolytic activity

cannot be determined based on our genomic data, but thus far our data serves as an interesting starting point for further studies of potential industrial enzymes in this species.

Production of secondary metabolites

We investigated two different aspects of the secondary metabolism in *P. arizonense*: Firstly, the genomic potential for secondary metabolite biosynthesis was assessed and 62 BGCs were identified using antiSMASH (Weber et al. 2015). Compared to nine other *Penicillium* species, *P. arizonense* proved to contain the highest number of PKS BGCs (28 in total). In a later paper, we further found that *P. arizonense* contained the highest number of PKS BGCs out of 24 different *Penicillium* species analyzed (**Paper III**). Secondly, we used Liquid Chromatography-Mass Spectrometry (LC-MS) to identify secreted secondary metabolites in the crude extract of *P. arizonense* cultivated on three different solid media known to induce secondary metabolite production (CYA, YES and OAT) (Frisvad 2012). A total of seven different compounds, or families of compounds, were detected in the media (**Table 1**), and many of these have potential medical applications as discussed in the original paper (Grijseels et al. 2016). Four of the detected compound families, austalides, pyripyropenes, tryptoquialanines and xanthoepocin, were not previously reported from species in section *Canescentia*. Many of the peaks in the chromatograms could not be mapped to a compound, thus highlighting the potential for discovery of novel secondary metabolites from *P. arizonense*.

Table 1. Detected compounds and BGCs in *P. arizonense*.

Detected compound or family of compounds	*P. arizonense* BGC with similarity to	No. of homologs[b]	Avg. similarity [%ID / %cov]
Austalides[a]	Mycophenolic acid (Regueira et al. 2011)	3/8	59/96
6-farnesyl-5-7-dihydroxy-4-methylphthalide[a]			
Pyripyropenes[a]	Pyripyropene (Itoh et al. 2010)	7/9	79/99
Tryptoquivalines[a]	Tryptoquialanine (Gao et al. 2011)	13/13	74/96
Fumagillin	Fumagillin-pseurotin (Wiemann et al. 2013)	16/16	82/95
Pseurotin A			
Xanthoepocin[a]	Aurofusarin (Frandsen et al. 2011)	7/11	49/96
Curvulinic acid	N/A	N/A	N/A

[a]Not previously seen in section *Canescentia*.
[b]Genes were considered homologs if: ID > 30% and coverage > 50%.

Linking secondary metabolites to BGCs

In order to identify which BGCs that were responsible for production of the detected compounds, we applied a BLAST based targeted approach. For each compound, we searched the literature to determine if the corresponding BGC had been characterized in another species (**Table 1**). This was the case for pyripyropenes (Itoh et al. 2010), fumagillin-pseurotin (Wiemann et al. 2013) and tryptoquivalines (Gao et al. 2011) (actually tryptoquialanines). The BGCs responsible for the biosynthesis of these compounds in other species showed in all cases high similarity to one of the detected

BGCs in *P. arizonense*, and the majority of gene cluster members were conserved as well (**Table 1**).

For the remaining compounds, we evaluated their chemical structure and biosynthesis, with respect to similarity to characterized pathways. Austalides contain a phtalide PK core which was detected in the crude extract of *P. arizonense* and also constitutes the PK core of mycophenolic acid (de Jesus et al. 1983). The mycophenolic acid BGC has been characterized in *P. brevicompactum* (Regueira et al. 2011), and one of the BGCs in *P. arizonense* showed homology to three of the genes. These genes correspond to the ones responsible for biosynthesis of the phtalide moiety of mycophenolic acid in *P. brevicompactum*, and the BGC could thus likely be responsible for austalide biosynthesis *P. arizonense*. Similarly, a probable BGC responsible for xanthoepocin biosynthesis in *P. arizonense*, was identified by having orthologs to seven genes in the BGC encoding the structurally related compound, aurofusarin, in *Fusarium graminearum* (Frandsen et al. 2011). For curvulinic acid, we were not able to identify characterized BGCs encoding similar products, and thus no BGCs could be confidently assigned to curvulinic acid production.

The putative austalide and xanthoepocin BGCs in *P. arizonense* showed lower sequence similarity to the most similar BGCs in other species compared to the cases were the same BGCs in another species was found (**Table 1**). This suggests a more distant evolutionary relationship, and demonstrates how the modularity of BGCs can be repurposed to evolve new pathways in fungi (Wisecaver and Rokas 2015). This modularity of the secondary metabolism provides potential for applying synthetic biology tools to develop new secondary metabolite pathways by recombining domains and modules to generate alternative products (Medema et al. 2012).

In summary, we have described a novel *Penicillium* species which possess promising potential for production of industrial enzymes and bioactive secondary metabolites with medical applications. Surprisingly it contained the most GH domains and PKS BGCs compared to related organisms. Our description of the species, combined with availability of its genome sequence provides a basis for further biotechnological exploitation of *P. arizonense*.

Paper III: Global biosynthetic potential of the *Penicillium* genus

To broaden the insights into to the potential of Penicillia for the biosynthesis of secondary metabolites, we decided to widen the scope and include the genomes of as many *Penicillium* species as possible. This included published genomes, as well as the ten sequenced in this PhD project, altogether 24 genomes of different *Penicillium* species. The resulting study (**Paper III**) constituted the first genus wide investigation of *Penicillium* genomics, and provided fundamental insights into the diversity of secondary metabolite pathways at the genus level.

Overview of *Penicillium* genomics

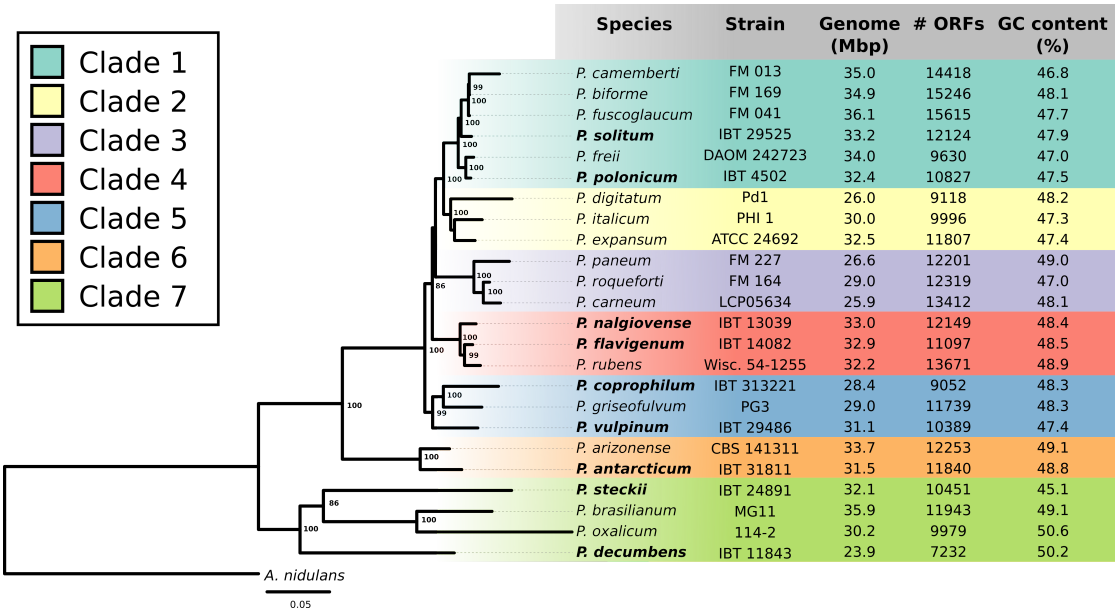
The first sequenced *Penicillium* genome was published in 2008 and encompassed the industrial penicillin producer *P. rubens* Wisconsin54-1255 (formerly *P. chrysogenum*) (van den Berg et al. 2008). Despite being published several years after the first *Neurosporra* and *Aspergillus* genomes in 2003 and 2005, respectively (Galagan et al. 2003; Galagan et al. 2005), it provided important insight into fungal biotechnology. The sequenced strain had been strongly exposed to classical strain improvement programs through random mutagenesis, and its genome sequence proved that this had led to amplification of the penicillin BGC and enhanced amino acid production (van den Berg et al. 2008). Four years later, the genome sequence of the postharvest pathogen *P. digitatum*, was published (Marcet-Houben et al. 2012), and this was followed by the genomes of other postharvest pathogens *P. expansum*, *P. italicum* (Ballester et al. 2015) and *P. griseofulvum* (Banani et al. 2016). These studies provided insights into the pathogenicity of *Penicillium* species, and their biosynthesis of mycotoxins such as patulin. For the purpose of studying production of lignocellulases, the genome of the industrial enzyme producers *P. oxalicum* was sequenced, and revealed a diverse set of plant cell wall degradation enzymes encoded (Liu et al. 2013b). Other sequencing efforts revolved around the importance of Penicillia for manufacturing food products, with the sequencing of *P. camemberti* and *P. roqueforti* (Cheeseman et al. 2014), and the year after an additional five *Penicillium* species important in the food industry were sequenced as well (Ropars et al. 2015). These studies demonstrated how HGT, among distantly related species, have contributed to adaptation to the environments found in cheeses.

The above walkthrough of *Penicillium* genomics, proves that Penicillia constitute a diverse genus with diverse industrial applications. However, one aspect that is conserved across these species, is the ability to produce secondary metabolites (Frisvad et al. 2004).

Penicillium phylogeny

In 2014 the *Penicillium* genus was defined to contain 354 accepted species (Visagie et al. 2014), and the phylogeny was recently revisited based on three marker genes (Houbraken et al. 2015). We conducted a more comprehensive assessment of the phylogeny of 24 *Penicillium* species using whole genome sequencing information to infer a maximum likelihood tree based on 1,389 single copy orthologous genes using the supermatrix approach (de Queiroz and Gatesy 2007) (**Figure 8**). The topology of the phylogram is in

agreement with the previous phylogenetic assessment based on marker genes (Houbraken et al. 2015). However, the quantitative nature of the branches in our genome based phylogenetic tree, constitute information on the relative phylogenetic distance (or time of divergence) between the species. For the further analysis, the 24 species were divided into seven clades corresponding to their specific section or subgenus as previously defined (Visagie et al. 2014).



Species	Strain	Genome (Mbp)	# ORFs	GC content (%)
P. camemberti	FM 013	35.0	14418	46.8
P. biforme	FM 169	34.9	15246	48.1
P. fuscoglaucum	FM 041	36.1	15615	47.7
P. solitum	IBT 29525	33.2	12124	47.9
P. freii	DAOM 242723	34.0	9630	47.0
P. polonicum	IBT 4502	32.4	10827	47.5
P. digitatum	Pd1	26.0	9118	48.2
P. italicum	PHI 1	30.0	9996	47.3
P. expansum	ATCC 24692	32.5	11807	47.4
P. paneum	FM 227	26.6	12201	49.0
P. roqueforti	FM 164	29.0	12319	47.0
P. carneum	LCP05634	25.9	13412	48.1
P. nalgiovense	IBT 13039	33.0	12149	48.4
P. flavigenum	IBT 14082	32.9	11097	48.5
P. rubens	Wisc. 54-1255	32.2	13671	48.9
P. coprophilum	IBT 313221	28.4	9052	48.3
P. griseofulvum	PG3	29.0	11739	48.3
P. vulpinum	IBT 29486	31.1	10389	47.4
P. arizonense	CBS 141311	33.7	12253	49.1
P. antarcticum	IBT 31811	31.5	11840	48.8
P. steckii	IBT 24891	32.1	10451	45.1
P. brasilianum	MG11	35.9	11943	49.1
P. oxalicum	114-2	30.2	9979	50.6
P. decumbens	IBT 11843	23.9	7232	50.2

Figure 8. Phylogenetic tree of the investigated *Penicillium* species. Maximum likelihood phylogenetic tree calculated using the supermatrix method based on 1,389 single copy core genes. *A. nidulans* was used as outgroup, bootstrap support is given from 100 bootstrap replicates, and the scale bar indicates the mean expected substitutions per site. Species marked in bold were sequenced in this study, and the clades correspond to section or sub-genus. Clade 1: section *Fasciculata*; clade 2: section *Robsamsonia*; clade 3: section *Roquefortorum*; clade 4: section *Chrysogena*; clade 5: section *Penicillium*; clade 6: section *Canescentia*; clade 7: sub-genus *Aspergilloides*.

Functional diversity of *Penicillium* species

Genes in the 24 species were grouped into orthologs using orthoMCL (Fischer et al. 2011). This enabled the identification of 3,248 gene families shared by all species, hence representing the core genome. In addition, 8,784 genes families were observed in a subset of species (at least two), thus representing the dispensable genome, and the pan-genome, here defined as the union of the two, was 12,032 gene families. In order to get an overview of the functional capabilities encoded in these different genome fractions, the genes were annotated using KOGs. We sorted the KOGs according to variance in the number of encoded genes within each KOG category, and found the greatest variation to be within secondary metabolism (Q), followed by the remaining metabolic subsystems: carbohydrate metabolism (G), amino acids metabolism (E) and lipid metabolism (I) (**Figure 9A)**.

Interestingly, the majority of the genes within secondary metabolism proved to be encoded in the core genome indicating that although there is great variation in the number of secondary metabolite biosynthetic genes, all the genomes did contain the different gene classes associated with secondary metabolism. The main variation in the number of genes

annotated as secondary metabolism per species was within cytochrome P450's. Although there is limited reporting of the functional capabilities of the pan- and core genome in fungi, it is generally assumed that the core genome contains essential and housekeeping genes (Vernikos et al. 2015). The fact that secondary metabolism is present in the core fraction of the *Penicillium* genomes seemingly opposes the notion of secondary metabolism being non-essential and an aspect of diversification. In our analysis, it was observed that the gene classes involved in secondary metabolite biosynthesis is present in all *Penicillium* species, but this doesn't necessarily mean that the pathways encoded in the genomes are the same (as shown in the next section). Another comparative genomics study on the genomes of 19 *Aspergillus* species was published two months prior to our *Penicillium* study, and the authors found a great diversity among several aspects of the genomes, including secondary metabolite biosynthesis pathways, although no analysis of the functional capabilities within the pan- and core genomes was conducted (de Vries et al. 2017).

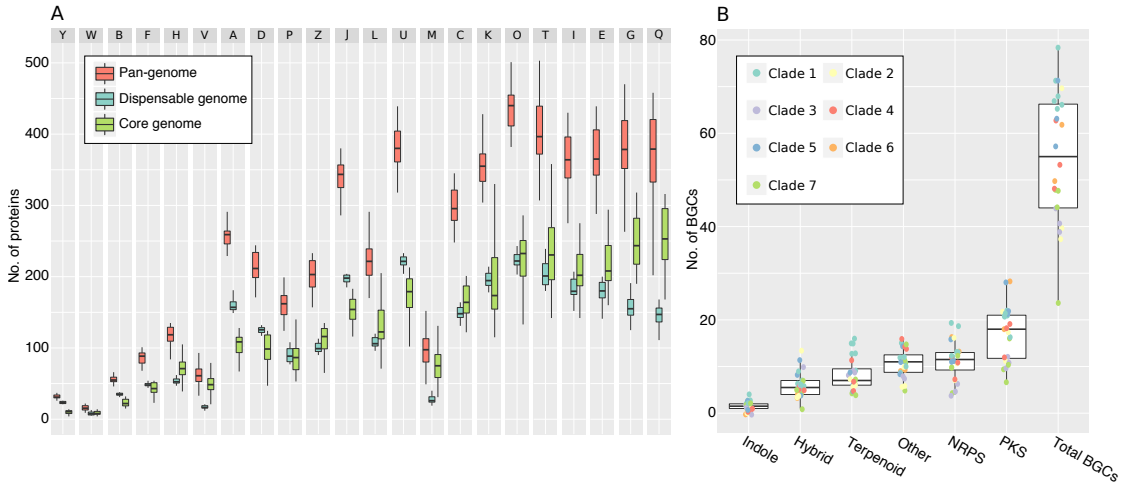


Figure 9. Functional analysis of the genomes of 24 *Penicillium* species. (A) Distribution of proteins allocated to different subsystems as defined by euKaryotic Orthologous Groups (KOGs). KOGs are sorted according to standard deviation in the pan-genome. KOG categories are as follows. For cellular processes and signalling, M is cell wall/membrane/envelope biogenesis, O is post-translational modification, protein turnover and chaperones, T is signal transduction mechanisms, U is intracellular trafficking, secretion and vesicular transport, V is defence mechanisms, W is extracellular structures, Y is nuclear structure, and Z is cytoskeleton. For information storage and processing, A is RNA processing and modification, B is chromatin structure and dynamics, J is translation, ribosomal structure and biogenesis, K is transcription, and L is replication, recombination and repair. For metabolism, C is energy production and conversion, D is cell cycle control, cell division and chromosome partitioning, E is amino-acid transport and metabolism, F is nucleotide transport and metabolism, G is carbohydrate transport and metabolism, H is coenzyme transport and metabolism, I is lipid transport and metabolism, P is inorganic ion transport and metabolism, and Q is secondary metabolites biosynthesis, transport and catabolism. (B) Distribution of BGC classes. Each boxplot represents the distribution of a class of BGCs as defined by antiSMASH. The class 'Other' contains BGCs that do not fit into any of the predefined categories of antiSMASH and rare BGC classes that were present only in few species.

Overview of the secondary metabolism

The main focus of this study was to evaluate the diversity and conservation in secondary metabolite biosynthesis pathways encoded in the species. Initially we mined the genomes for BGCs using antiSMASH, and could detect a total of 1,317 BGC in the 24 genomes (**Figure 9B).** Although, all genomes contained predicted BGCs there was a large variation between the species, with 78 predicted BGCs in *P. polonicum* and 22 in *P. decumbens*.

PKS and NRPS BGCs were the two major classes of BGCs detected (**Figure 9B)** in agreement with previous observations (Keller et al. 2005). On average we found 55 BGC per species, where recently an average of 41 BGC per species was found in 19 *Aspergillus* genomes (de Vries et al. 2017). This indicates a greater potential for secondary metabolite biosynthesis in *Penicillium* compared to *Aspergillus*, although different approaches for BGC detection were used, thus a direct comparison should be interpreted with caution.

The species from clade 1 contained more terpenoid BGCs than species from the other clades, and generally showed a high number of BGCs encoded in their genomes (**Figure 9B)**. It is interesting to speculate whether these species might be better equipped for production of terpenoids compared to other *Penicillium* species. Besides clade 1, there were no clear correlations between phylogeny and specific BGC classes (**Figure 9B)**.

Clustering of BGCs

To assess the conservation of secondary metabolite biosynthetic pathways across species, we grouped the detected BGCs into gene cluster families (GCFs). Since PKS and NRPS BGCs were the most abundant classes found in the genomes, we decided to focus on these for the clustering. We used the KS domain of PKSs and the C domain of NRPSs to assess the similarity between BGCs as previously discussed in **Paper I**.

Based on pair-wise alignments of all KS domains, a histogram of the pairwise sequence identities revealed a bimodal distribution, were the majority of the domains showed 20-60% identity, while a second group of KS domains showed higher similarity. From this, we selected a threshold of 74% identity, to separate these two distributions (**Figure 10A**). A similar bimodal distribution was observed based on pair-wise alignment of the C domains, although they generally were less similar, thus resulting in a lower threshold of 64% identity (**Figure 10B**). The selected thresholds were considerably lower than those chosen by Ziemert et al. (2014), where 90% and 85% identity was used as thresholds for grouping KS and C domains, respectively, and by extension, PKSs and NRPSs. In that analysis however, different isolates of only three different species *Salinispora* species were compared. Our lower thresholds correspond to a greater phylogenetic diversity in our dataset of 24 *Penicillium* species.

Based on the selected domain thresholds, PKS and NRPS BGCs were grouped into GCFs. To evaluate the quality of the clustering, a manual assessment of the synteny between clustered BGCs confirmed a good conservation of the sequences (**Figure 11B,C**). Following this grouping, we annotated the identified BGCs relative to the MIBiG database (Medema et al. 2015), which contains information on biosynthetic loci that have been linked to a compound. We applied the same approach and the same thresholds to link BGCs from the MIBiG database to the *Penicillium* BGCs, as we did when grouping the *Penicillium* BGCs based on KS and C domains. This allowed us to annotate 127 of the 798 *Penicillium* PKS and NRPS BGCs corresponding to 16%, to a pathway and a metabolic end-product (**Figure 10C,D**). In order to confirm these findings, we cross-referenced these 127 predictions and could validate 87 of the species:metabolite associations to the exact or a related compound from published literature, or based on chemical analysis conducted in the study. The remaining predicted species:metabolite relationships, either constitute new producers of known compounds, or false predictions that might encode pathways that are similar to the predicted ones.

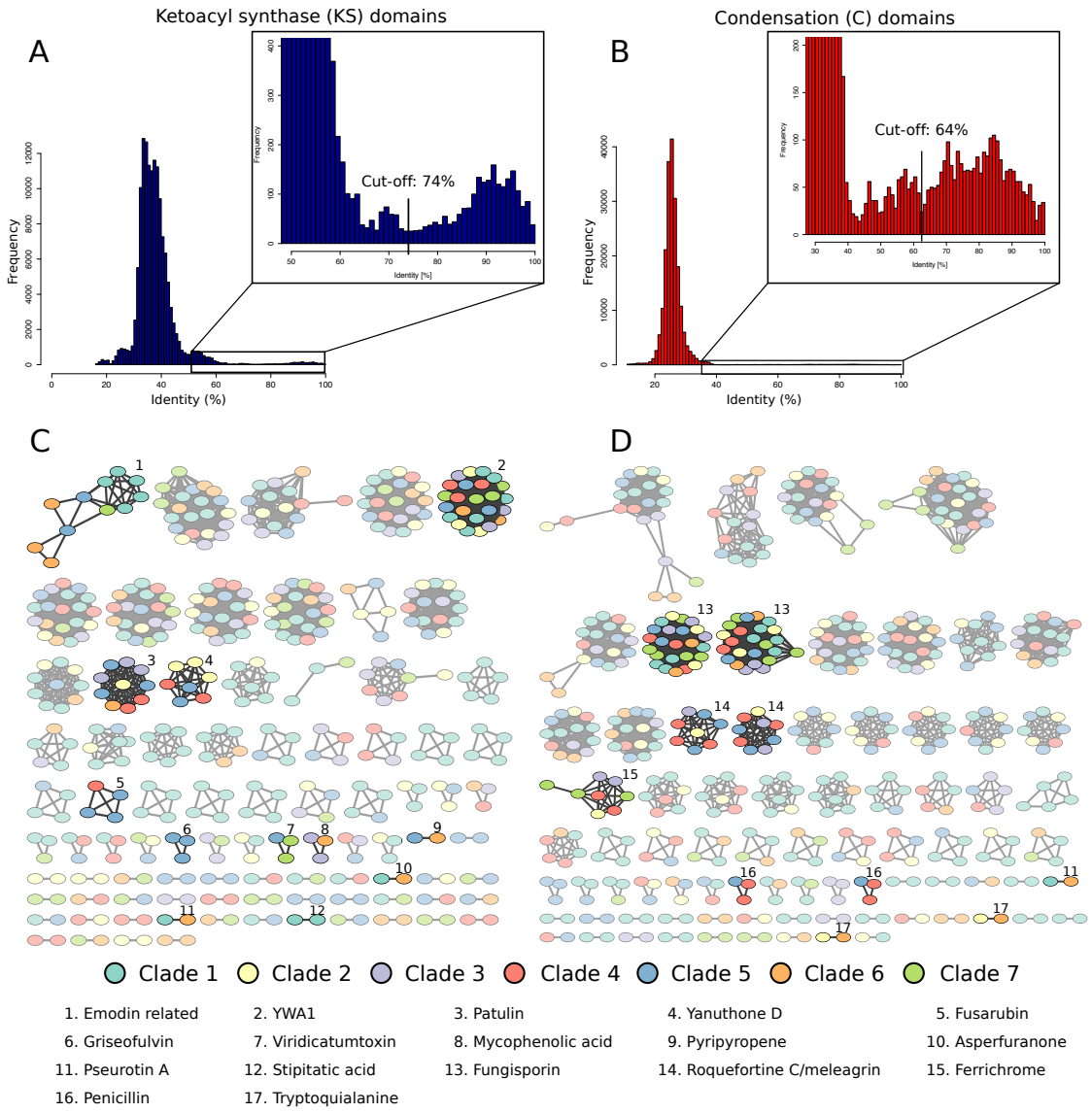


Figure 10. Global overview of biosynthetic gene cluster (BGC) similarity in 24 *Penicillium* genomes. Histograms of the similarity of (A) KS domains identified in PKS BGC and (B) C domains identified in NRPS BGCs of *Penicillium* species. Each histogram indicates the threshold for grouping the corresponding BGCs into GCFs. Network representation of GCFs based on (C) KS domains for PKS BGCs and (D) c domains for NRPS BGCs. Each node represents a domain and by extension a BGC, and the edges connect domains that are similar. Nodes are colored according to the clade of the species and highlighted clusters represent GCFs that mapped to an entry in the MIBiG database (Medema et al. 2015).

Production of yanuthones in *Penicillium*

As a proof of concept, we decided to look further into the annotated yanuthone encoding BGCs that were predicted to be present in seven of the species (**Figure 10C**). Previously, only one intermediate of the yanuthone pathway, 7-deacetoxyyanuthone, has been reported from *P. chrysogenum* (Maskey et al. 2005), as well as from a marine *Penicillium* species (Li et al. 2003), so we wanted to assess if the Penicillia also could synthesize the end-product of the *Aspergillus* version of the pathway, yanuthone D. The biosynthesis of yanuthone D has previously been characterized in *A. niger* (Holm et al. 2014), and yanuthones are of medical interest as they have been shown to exhibit antifungal activity (Holm et al. 2014; Petersen et al. 2015). Interestingly, the identified *Penicillium* version

of the BGC contained one additional conserved gene, which is not present in *A. niger*. This additional gene did not show any obvious hits in the NCBI nr-database, although a membrane bound prenyl cyclase from the pyripyropene BGC in *A. fumigatus* showed 32% identity. It seems plausible that the BGC could contain a prenyl cyclase since yanuthones contain a prenyl moiety that could undergo cyclization, similarly to what is seen in the pyripyropene biosynthesis (Itoh et al. 2010).

To assess production of yanuthones, the seven *Penicillium* species were grown on different solid media, and crude extracts were analyzed using LC-MS. Using a reference standard, yanuthone D and E were detected in *P. rubens*, while only yanuthone E was detected in *P. flavigenum*. We further searched for yanuthone isomers generating fragment ions, as the known standards, and could identify a new yanuthone derivative. These findings validate our prediction approach and shows how it can be applied in predicting production of known and novel compounds in new species. The production of a novel yanuthone further suggests the participation of additional enzymatic activities in the *Penicillium* version of the yanuthone BGCs, that could originate from the additional gene identified in the cluster. The exact function of the additional *Penicillium* gene is however, still dubious since the prenyl moiety of the novel yanuthone, was unchanged.

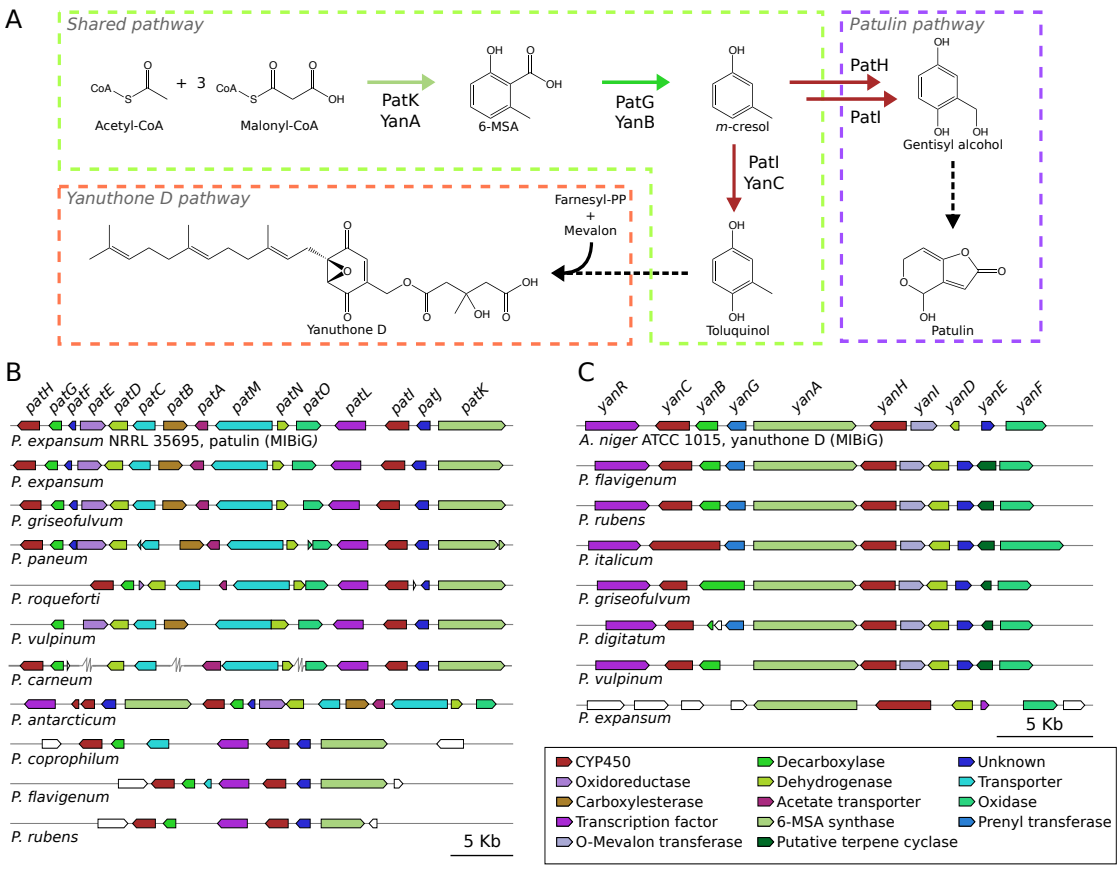


Figure 11. Overview of 6-MSA pathways in *Penicillium* species. (A) Biosynthetic pathways of patulin and yanuthone D. Dashed arrows indicate multiple steps. (B) Predicted patulin BGCs relative to the version in *P. expansum* (Tannous et al. 2014) (C) Predicted yanuthone D BGCs relative to the version in *A. niger* (Holm et al. 2014).

Yanuthones were not produced in the remaining five *Penicillium* species in the conditions tested. In *A. niger*, yanuthones have proved to be very specific and only induced on YES

medium where it is represented as little peaks in the chromatogram (Holm et al. 2014), while in *P. chrysogenum* PDA medium induced yanuthones (Maskey et al. 2005). We cultivated all seven *Penicillium* species on solid YES and PDA, and yanuthones were observed for *P. rubens* and *P. flavigenum*, only on PDA, and still as minor peaks. Possibly other conditions are required for induction of yanuthones in the five remaining Penicillia.

Evolution of *Penicillium* BGCs

In addition to providing information about the biosynthetic capabilities of *Penicillium* species, our analysis further enabled insight into the evolution of fungal BGCs. In particular, we identified two groups of BGCs containing 6-MSA synthases (6-MSAS), annotated as synthesizing the 6-MSA based compounds yanuthones and patulin (**Figure 11B,C**). Examining the prevalence of these 6-MSAS based pathways in *Penicillium* proved that no species from clade 1 (section *Fasciculata*) and clade 7 (subgenus *Aspergilloides*) contained the BGCs, and suggests that 6-MSAS could have been lost by their divergence (**Figure 12A**). In contrast, 6-MSAS was present in most other species. We also observed that the yanuthone producing species (*P. rubens* and *P. flavigenum*), does not also produce patulin (Frisvad et al. 2004), but rather contain fossil patulin BGCs that are likely not functional. This suggest that yanuthone and patulin production does not commonly co-occur in one species, possibly to avoid cross chemistry, since their pathways are highly redundant (**Figure 11A**).

Among the annotated patulin BGCs, we noted that *P. antarcticum* had a new conformation that is distinct from previously identified ones in *P. expansum* and *A. clavatus* (Artigot et al. 2009) (**Figure 12B**). We found that the version of the BGC in *P. antarcticum* served as an intermediate between the two previously identified, with the dominant architecture in *Penicillium* being the one present in *P. expansum*. This is in agreement with the phylogeny of the species, where *P. antarcticum* is the earliest diverging species, among the ones containing patulin BGCs. This provides insight into the rate of which rearrangements and thus diversification of BGCs occur in fungi.

This concludes part I of my thesis describing the large biosynthetic potential of filamentous fungi for the production of future drug leads. This work can be exploited for the discovery of new producers of known secondary metabolites or new derivatives of secondary metabolites, by searching for BGCs with extra genes. In order to develop a feasible production processes of new secondary metabolites, it is valuable to have a thorough understanding of the physiology that governs secondary metabolite production. In part II, I will describe studies on how to translate the identification of novel secondary metabolites into industrial production processes.

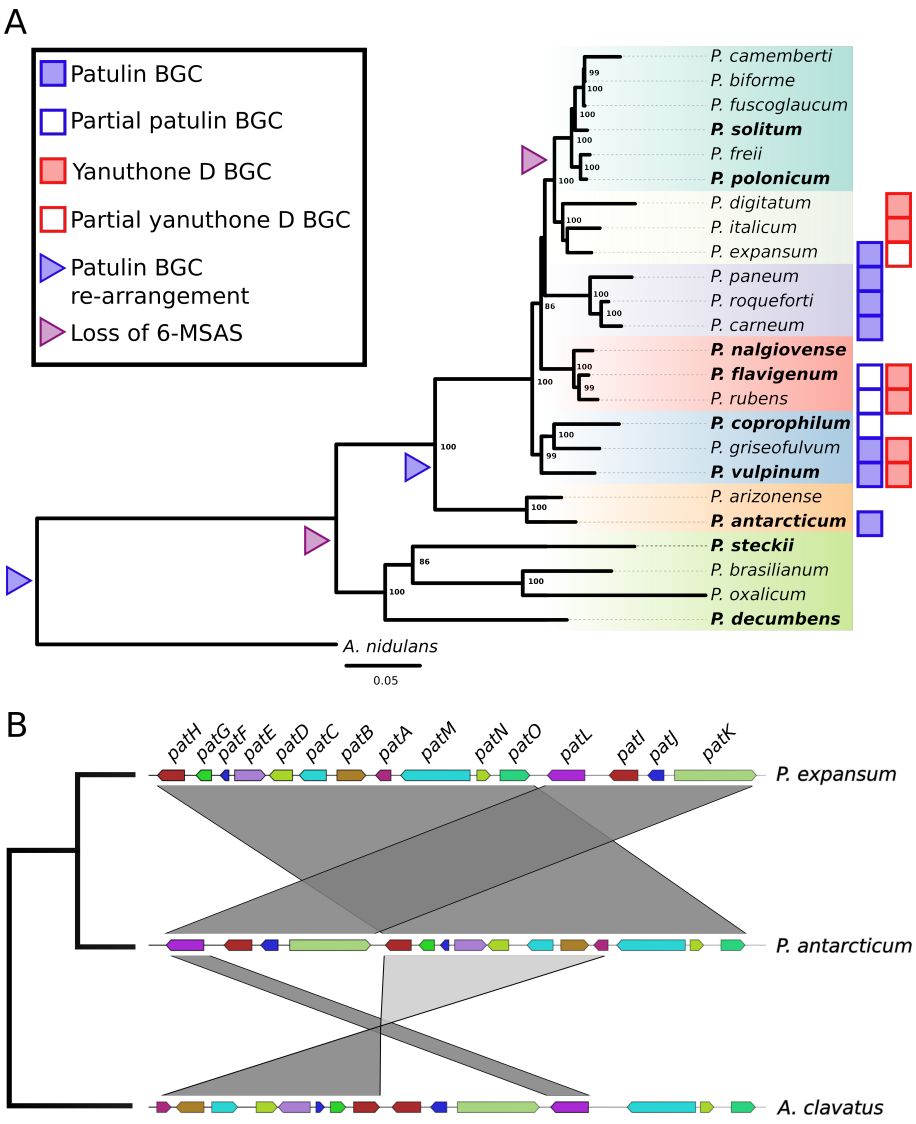


Figure 12. Evolutionary events of 6-MSA pathways in *Penicillium*. (A) Overview of evolutionary trajectory of 6-MSA based pathways in *Penicillium* species. (B) Gene organization of patulin BGC in *P. antarcticum*, discovered in this study compared to two previously published patulin BGCs in *P. expansum* (Tannous et al. 2014) and *A. clavatus* (Artigot et al. 2009).

Part II: Characterizing secondary metabolism using systems biology

Once a promising compound of interest has been identified, the next step is to establish a feasible production process. Systems biology can aid in characterizing cell factories and provide insight into how a strain and overall process economics can be optimized for production.

The traditional approach to establish production processes using cell factories has been to select naturally high yielding isolates producing the compound of interest. This has historically been a successful approach in many industrial production processes of chemicals such as glutamate production by *Corynebacterium glutamicum* (Kinoshita et al. 1957) and citric acid production by *A. niger* (Currie 1917). For large-scale production of penicillin, various *Penicillium* species and isolates were screened for production. This led to the discovery of a high yielding strain, *P. rubens* NRRL 1951, that can be traced back to a moldy cantaloupe melon from the food market in Peoria, Illinois, USA. The isolate, is the ancestor of all current industrial penicillin production, and proved to possess a major potential for further optimization. Following decades of classical strain improvement programs through random mutagenesis, penicillin production was increased more than 10,000 fold (Thykaer and Nielsen 2003).

Nowadays, an increasingly popular strategy is to transfer the pathway of interest to a platform cell factory which has been optimized for industrial production. Such organisms are advantageous since they are well characterized, and a number of gene editing and gene expression tools are established (Nielsen and Keasling 2016). In addition, existing systems biology tools such as GEMs, can be exploited to easily understand the context of expression of a heterologous pathway. The yeast, *S. cerevisiae* serves as an attractive platform for secondary metabolite production and has been successfully used to express heterologous pathways of fungal PKs (Rugbjerg et al. 2013) and NRPs (Awan et al. 2017). Examples of heterologous production, however, is limited in the scientific literature, and often associated with a low yields (Awan et al. 2017). Several factors could influence why it is troublesome to express PKS and NRPS pathways in yeast, but a general limitation is that yeast does not natively produce secondary metabolites, and thus might lack some fundamental cellular features for efficient production. This could include correct compartmentalization, since many secondary metabolite biosynthesis pathways have proved to utilize complex multistep compartmentalization processes (Roze et al. 2011; Kistler et al. 2015). Further, correct folding of the large PKS and NRPS enzymes might not always take place in heterologous hosts such as yeast (Siewers et al. 2009). For these reasons, we decided to evaluate the potential of wild type *Penicillium* species as native cell factories for secondary metabolite production.

Paper IV and V: *Penicillium* species as cell factories

The ten *Penicillium* species that were genome sequenced in **Paper II** and **III** showed a promising potential for biosynthesis of secondary metabolites. We decided to further investigate if these species also had potential as cell factories. All species were cultivated in controlled bioreactor fermentations and characterized at the physiological level for growth and production of secondary metabolites (**Paper IV**). For six of the species, we further characterized them at the transcriptional level (**Paper V**).

Physiological characterization of *Penicillium* species

Cultivations were carried out in submerged conditions in controlled bioreactors in two different media: One defined medium (DM) for *Penicillium* based on glucose and ammonium, and one complex medium (CM) called CYA, based on yeast extract, sucrose and nitrate. The DM was selected to characterize the physiology of the species, while the complex medium was selected because it has been reported to induce production of secondary metabolites (Frisvad 2012) and is of industrial relevance. During the cultivations, we monitored the morphology of the species, biomass dry weight, CO_2 exhaust, glucose consumption and secretion of secondary metabolites (**Figure 13**).

Overall, we observed a large difference in growth in the two media (**Table 2**). In DM, six of the species grew as pellets or clumbs, while *P. nalgiovense*, *P. decumbens*, *P. antarcticum* and *P. arizonense* had dispersed mycelium in the fermenters. In contrast, all species grew with dispersed mycelium in the CM. Growth rates of all species were in the range of 0.14-0.22 h^{-1} in DM and 0.17-0.29 h^{-1} in CM, and the biomass yields were in the range of 0.25-0.67 g DW g^{-1} glucose in DM. These results are comparable to what has been achieved with an industrial penicillin *P. rubens* strain (Robin et al. 2001), thus emphasizing the potential of using *Penicillium* biodiversity in industrial fermentations in terms of growth rate and biomass yield. Interestingly, the two species that grew the fastest in the CM, were also the species with the smallest genomes. Genome minimization could be an evolutionary mechanism by these species to optimize growth rate, by reducing the costs of genome replication. The DM in this study was based on a previously published medium for *P. rubens* (Thykaer et al. 2008), and adapting the medium to the individual *Penicillium* species could potentially further improve the morphology and growth rates, especially since the medium is optimized for penicillin production and not for growth specifically. Although dispersed mycelium resulted in the highest growth rates, the desirable morphology depends on the production process in question, as morphology is known to strongly affect production of metabolites and enzymes (Krijgsheld et al. 2013). For examples, for optimal production of citric acid in *A. niger*, pellet formation is desirable (Gomez et al. 1988).

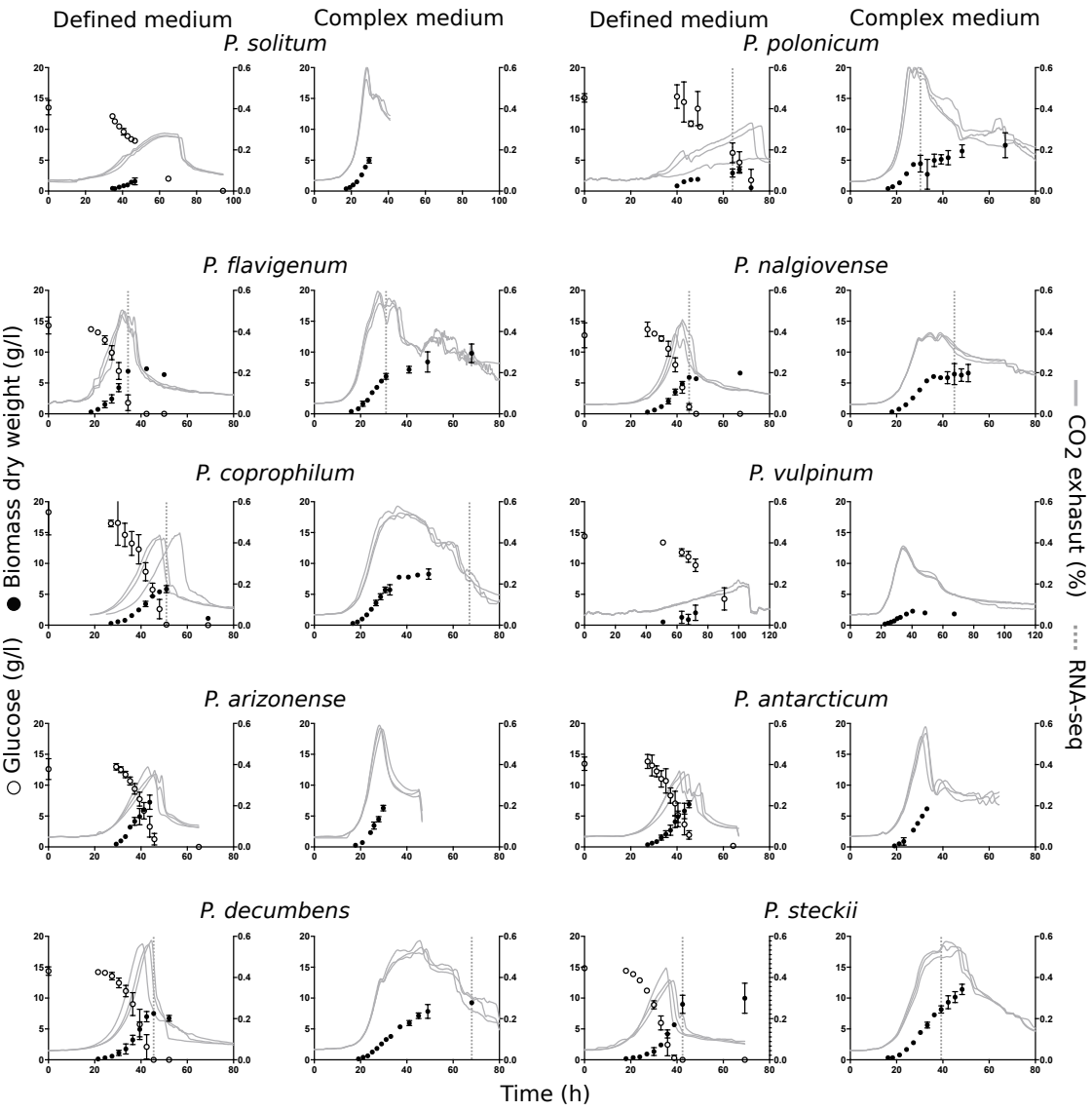


Figure 13 Growth profiles of *Penicillium* species. Ten *Penicillium* species cultivated in controlled bioreactor fermentations in a defined medium for *Penicillium* and a complex medium. All cultivations were conducted in biological triplicates.

Table 2. Physiological parameters of *Penicillium* species. Growth and morphology of ten *Penicillium* species relative to an industrial isolate of penicillin producing *P. rubens*.

	Defined medium			Complex medium		
Species	**Y_{sx}**	**μ_{max}**	**Morphology[a]**	**μ_{max}**	**Morphology[a]**	**Reference**
P. solitum	0.29±0.13	0.14±0.02	Pellets/clumbs	0.22±0.01	Dispersed	This study
P. polonicum	0.25±0.01	n/a	Dispersed	0.22±0.01	Dispersed	This study
P. nalgiovense	0.60±0.17	0.21±0.01	Dispersed	0.21±0.01	Dispersed	This study
P. flavigenum	0.58±0.05	0.22±0.01	Pellets/clumbs	0.26±0.00	Dispersed	This study
P. coprophilum	0.37±0.03	0.17±0.02	Pellets/clumbs	0.28±0.01	Dispersed	This study
P. vulpinum	0.54±0.14	n/a	Pellets/clumbs	0.21±0.01	Dispersed	This study
P. arizonense	0.67±0.04	0.14±0.01	Dispersed	0.24±0.00	Dispersed	This study
P. antarcticum	0.58±0.03	0.19±0.00	Dispersed	0.17±0.03	Dispersed	This study
P. steckii	0.37±0.02	0.15±0.02	Pellets/clumbs	0.21±0.01	Dispersed	This study
P. decumbens	0.55±0.02	0.19±0.02	Dispersed	0.29±0.02	Dispersed	This study
P. rubens	0.45	0.19	N/A	N/A	N/A	(Robin et al. 2001)

± denotes standard deviation where $n = 3$.
[a] morphology of the mycelium in the submerged cultivations.

Secondary metabolite production

The fermentation media were analyzed for secreted secondary metabolites using LC-MS, and intermediates or end-products of 14 different secondary metabolite pathways were detected in the fermentation media (**Figure 14**). In addition, a number of unidentified peaks were detected. Surprisingly, for the majority of compounds no differential effect of the two different fermentation conditions were observed. The compounds that were differentially produced in the two media were fungisporin by *P. coprophilum*, penicillic acid and verrucofortine by *P. polonicum* and atlantinone by *P. solitum*. Possibly, the submerged condition was a more important environmental factor determining the induction of BGCs than the differences in nutrient composition in the two media. In comparison, we detected eight secondary metabolites produced by *P. arizonense* on solid CYA medium (**Paper II**), while in submerged cultivations in the same medium (CM), only pseurotin was identified. Overall, this demonstrates that submerged cultivations are not a good approach for screening the diversity in secreted secondary metabolites. On the other hand, it might be a useful strategy for achieving high production levels since competition from other secondary metabolite pathways is minimized. Disruption of competing secondary metabolite pathways has been shown to increase production of certain secondary metabolites and enable production of new ones in fungi (Salo et al. 2015). This has also been exploited to increase production of specific secondary metabolites in *Actinobacteria* (Komatsu et al. 2010; Gomez-Escribano and Bibb 2011).

The species with a high growth rate on the CM and production of few secondary metabolites such as *P. flavigenum* might thus contain a great potential for further industrial development. Detection of secondary metabolites was not quantitative, but it would be interesting to investigate the titres of secondary metabolites in submerged fermentations of *Penicillium* species.

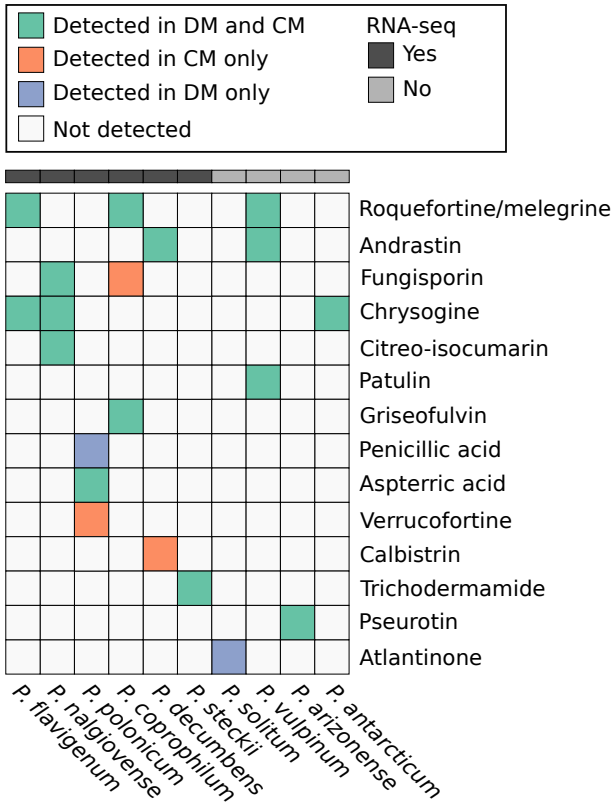


Figure 14. Secondary metabolite families produced during bioreactor batch fermentations. Compounds were detected by LC-MS/MS based on fragmentation pattern and UV/vis spectra of reference standards.

Transcriptional landscape of *Penicillium*

To further gain insights into the regulations of secondary metabolite production and to assess the transcriptional profile during secondary metabolite biosynthesis, we selected six of the ten species for transcriptome analysis. The species were chosen to represent the phylogenetic diversity of the *Penicillium* genus (**Figure 15A**). Samples for transcriptome analysis were collected in the stationary phase, several hours after the CO_2 exhaust had peaked, indicating that a nutrient had been depleted (**Figure 13**). The time point was selected in order to ensure activity of the secondary metabolism, since many secondary metabolite pathways are induced at stress conditions, such as nutrient depletion (Brakhage 2013).

In order to investigate the conserved transcriptional responses across species, we initially identified orthologous genes in the six genomes. We found 4,296 core gene families (in comparison we found 3,248 core gene families in the 24 *Penicillium* species in **Paper III**), and among these, 3,782 gene families were present only in a single copy in each genome (**Figure 15C**). These single copy core genes were important for the further comparative transcriptome analysis as they enabled a direct comparison of gene expression between species, while the remaining core genes include duplications and hence cause problems in identifying which paralogs correspond to each other in different genomes.

GEMs were reconstructed (further described in **Paper VI**) and used as a framework for annotation of metabolism and as a roadmap of the metabolic pathways and capabilities

of the species. We found that 582 single copy core genes were part of metabolism, and that 1220 metabolic reactions in the GEMs were catalyzed by the corresponding enzymes (**Figure 15D**). These 1220 core reactions were significantly depleted for reactions involved in secondary metabolism, in particular alkaloid and terpenoid biosynthesis pathways.

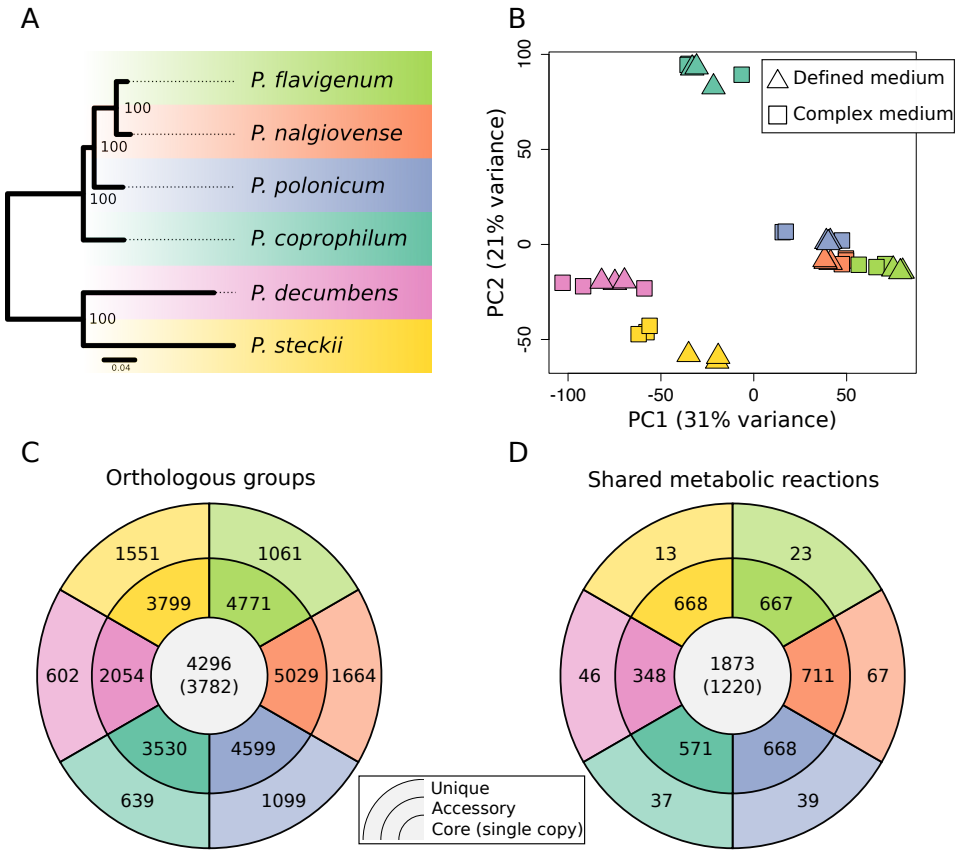


Figure 15. Overview of differences and similarities in six *Penicillium* species. (A) Phylogenetic tree of the species used in this study. (B) Principal component analysis of the six *Penicillium* species based on the normalized expression of single copy core genes. (C) Number of genes and (D) reactions within each of the species that are shared by all species (core), shared by a subset of species (accessory) or specific to one species (unique).

Based on the expression levels of the single copy core genes, a principal component analysis (PCA) was conducted (**Figure 15B**). The individual samples grouped according to the phylogeny of the species, while medium had little effect on the global gene expression pattern. This suggests that the gene regulation of core genes might be strongly dependent on the evolutionary distance between species. In yeast, evolutionary distance has proved to be correlated to differences in gene regulation. In particular, it has been found that gene duplication events strongly increase divergence in gene regulation (Thompson et al. 2013). Our results indicate that these regulatory differences go beyond gene duplication, as we only focused on genes without paralogs (the singly copy core genes). Further, in the next study (**Paper IV**), we show that the metabolic capabilities are largely conserved across *Penicillium* species, thus suggesting that regulatory differences could give rise to the metabolic diversity of the species, by inducing/repressing pathways.

Differentially expressed genes (DEGs) were identified (adjusted $p<0.05$), using the DESeq2 procedure (Love et al. 2014). DEGs were defined as up or down-regulated in the CM with the DM as reference. We observed a large variation in the number of affected

genes in the species, with *P. steckii* showing the greatest difference between the two media (up-regulated: 1,885; down-regulated: 2,363) while *P. decumbens* had the fewest affected genes (up-regulated: 327; down-regulated: 331) (**Figure 16A**).

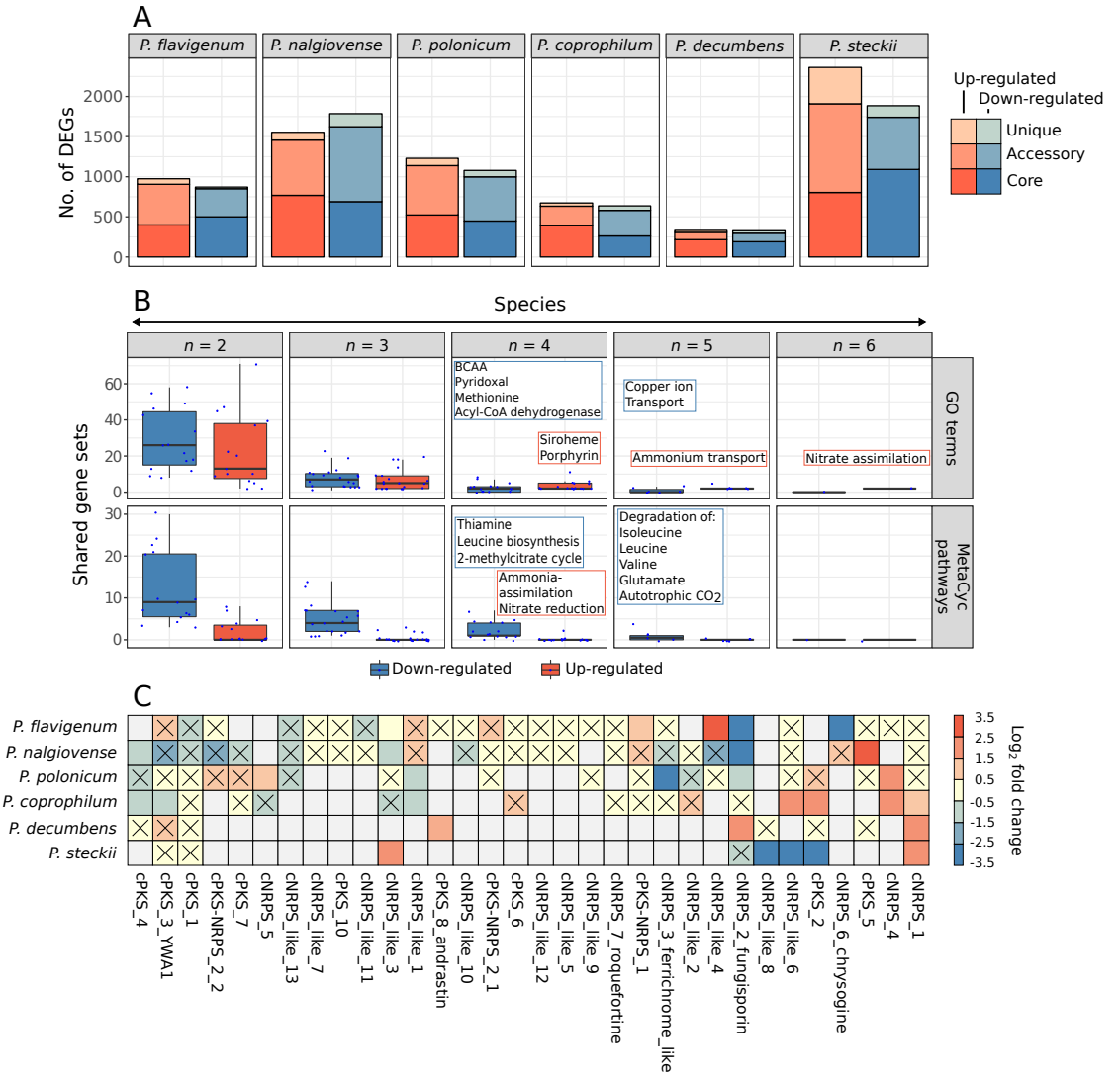


Figure 16. Overview of main transcriptional changes. (A) Number of differentially expressed genes in complex versus defined medium. (B) Gene set enrichment analysis identifying cellular processes shared by at least two species. Each data point in the box-plots represent a gene set enriched by a combination of *n* species and grouped as either up or down-regulated. The main shared processes are written in frames. (C) Effect of media on the expression of secondary metabolite backbone genes that were present in at least two species. Cells in the heatmap marked with a cross represents backbone genes that were not significant differentially expressed. Up and down-regulation refers to expression levels in complex medium compared to defined medium. Grey cells indicate backbone genes not detected in the species.

In order to identify the function of these genes, we used a GO term annotation of the genomes and a MetaCyc annotation of metabolic genes based on the reconstructed GEMs. Among the DEGs, we found two genes that were up-regulated in all six species and these were annotated as a nitrate reductase (ortholog *NIAD* in *A. nidulans*) and an ammonium uptake transporter (ortholog *MEAA* in *A. nidulans*). Both genes have been shown to respond to nitrate availability: *NIAD* reduce nitrate to nitrite intracellularly and is known to be up-regulated in response to hypoxia (Terabayashi et al. 2012), while *MEAA* is a low affinity ammonium transporter which has proved to be up-regulated under nitrogen

starvation and induced by nitrate (Schinko et al. 2010). No shared DEGs were down-regulated in all species.

Although only two genes were differentially expressed in all six species, we conducted a gene set analysis to investigate if specific GO terms or MetaCyc pathways were enriched across multiple species. In accordance with the shared DEGs, we observed only one gene set to be enriched across all six species and this was an up-regulation of the nitrate assimilation GO term (**Figure 16B**). Also, ammonium transport and ammonia assimilation was up-regulated in five and four species for GO terms and MetaCyc pathways, respectively. The GO term, copper ion transport, was down-regulated in five species. Among the MetaCyc pathways, degradation of a number of amino acids was observed, in particular the BCAAs and glutamate. These pathways were down-regulated in all species except *P. coprophilum*.

Taken together, remarkably few gene sets were enriched across a majority of the species. These results indicate that although the species are within the same genus, their responses to the different fermentation conditions were highly diverse. Thus, future work should be careful when extrapolating information between species from a diverse genus such as *Penicillium*.

Expression of biosynthetic gene clusters

In the six genomes, we identified a total of 311 BGCs, that we grouped into 42 GCFs consisting of BGCs from at least two species. Among these GCFs, 32 contained backbone genes of the classes PKS, NRPS, or PKS-NRPS, and seven of these could be linked to a pathway. We applied the BIG-SCAPE algorithm (Navarro-Muñoz, Yeong, Medema et al., in preparation) to group and annotate BGCs relative to the MIBiG database, and could connect six GCFs to a pathway. The BIG-SCAPE algorithm uses a combination of parameters including order and similarity in shared protein domains, and overall the resulting GCFs were very similar to the ones obtained in **Paper III**. Based on the PKS and NRPS genes, the expression of BGCs within the 32 GCFs was evaluated (**Figure 16C**). An equal distribution between up and down-regulated BGCs confirmed that the media resulted in no major differences in the number of induced BGCs, in accordance with the observations from the analytical detection of secondary metabolites (**Figure 14**).

We correlated our BGC annotation with the detected secondary metabolites in the fermentation media (**Figure 14**). In four of six cases, we detected the secondary metabolites corresponding to the annotated BGCs. This included: (*i*) andrastins by *P. decumbens*, (*ii*) chrysogines by *P. flavigenum* and *P. nalgiovense*, (*iii*) roquefortine/meleagrin intermediates by *P. coprophilum* and *P. flavigenum* (*iv*) and fungisporin by *P. coprophilum*, *P. nalgiovense* and *P. flavigenum*. We further looked into the expression of the gene members, based on previous characterization of the annotated andrastin BGC in *P. roqueforti* (Rojas-Aedo et al. 2017) and chrysogine BGC in *P. rubens* (Viggiano et al. 2017) (**Figure 17**). All genes in the andrastin BGC in *P. decumbens* were up-regulated, and oppositely chysogine genes in *P. flavigenum* were down-regulated. For the chrysogine BGC in *P. nalgiovense*, however, some genes were up-regulated while others down-regulated. Specifically, genes *chyE* and *chyH* were significantly down regulated, while *chyD* and *chyA* were up-regulated (**Figure 17B**). ChyA and ChyD catalyze the two first steps in the pathway, while ChyE and ChyH are thought to catalyze later steps and the differences in expression could thus constitute a

temporal transcriptional control based on when the individual enzymes are needed in the pathway.

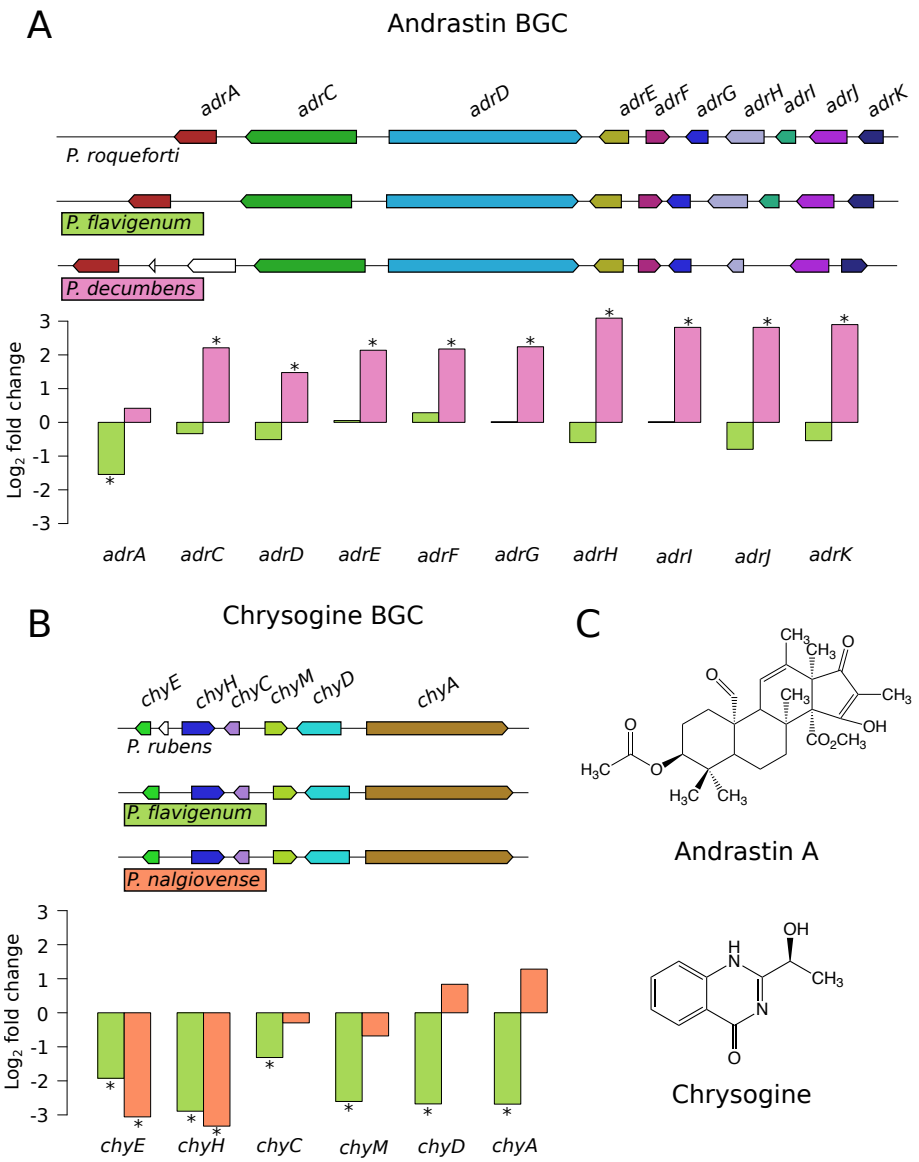


Figure 17. Expression of BGCs in *Penicillium*. (A) andrastin BGC, (B) chrysogine BGC and (C) chemical structure of the end-products of the pathways. Barplots represent the $\log_2$ fold change in expression level in complex medium relative to defined medium. Asterisk (*) denotes genes that were significantly differentially expressed.

Identification of coexpression modules

We showed above that few DEGs were shared among the species and the gene sets that were enriched proved to be present in only a subset of the species (**Figure 16A,B**). In order to further identify if there were certain groups of genes which expression was correlated to each other, we conducted an analysis of coexpression across the six species. For this, the Pearson correlation coefficient (PCC) was computed among the 3,782 single copy core genes, as well as the 33 PKS and NRPS genes identified in the GCFs. The pairwise correlation in expression levels among these 3,815 genes constituted a weighted

coexpression network. This network served as a starting point to identify modules of highly coexpressed genes (**Figure 18**). The coexpression network was decomposed into nine subnetworks by removing correlations (edges) between genes (nodes) by applying varying PCC cut-off's. Subsequently, modules of correlated genes were identified in the subnetworks using the clusterONE algorithm (Nepusz et al. 2012), and a total of 54 modules were found after removing redundant ones.

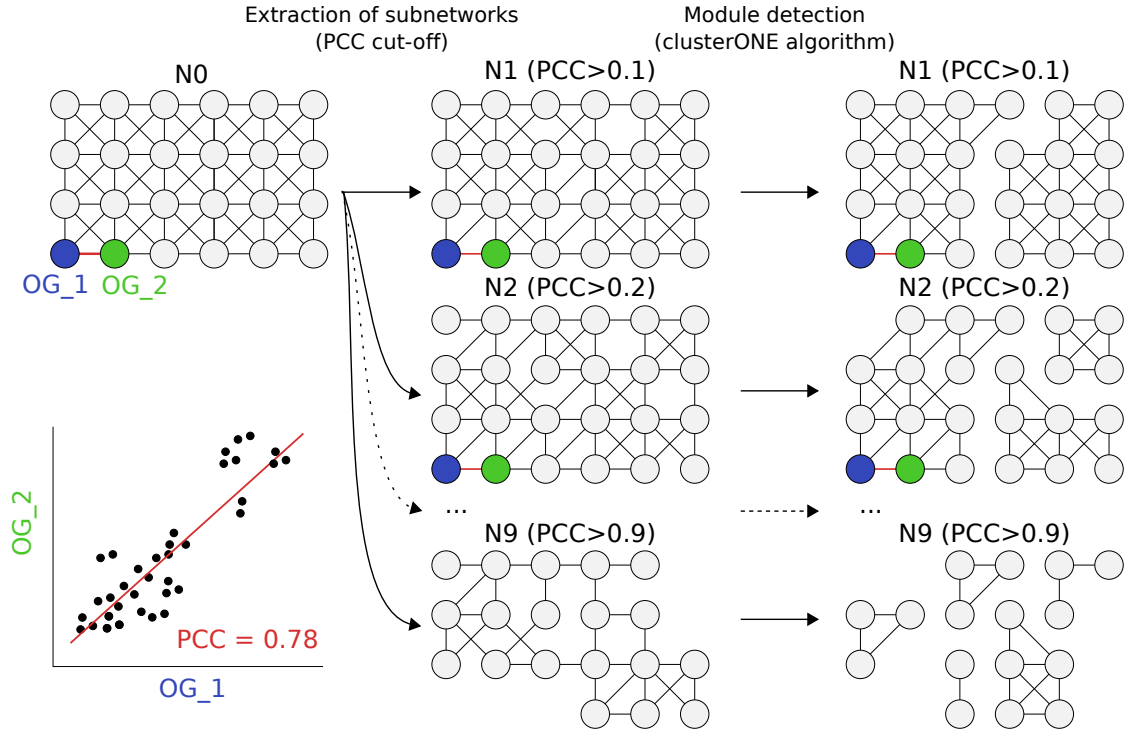


Figure 18. Overview of coexpression network analysis. Network "N0" is the global coexpression network where each node represents a group of orthologous genes, and edges represent Pearson's correlation coefficient (PCC) weights. Nine subnetworks N1-N9 were generated by removing edges based on varying PCC cut-off's. For each of these coexpression subnetworks, modules of highly coexpressed genes were detected using the clusterONE algorithm (Nepusz et al. 2012).

For these 54 modules of coexpressed genes, an enrichment analysis (hypergeometric test) was conducted based on the MetaCyc annotation of metabolism, and a total of 29 modules were enriched for at least one pathway (**Figure 19A**). The pathways that were enriched in the most modules were related to proteinogenic amino acid metabolism, both degradation and biosynthesis. A total of nine modules were enriched for degradation of amino acids, in particular degradation of leucine, valine and tyrosine. Conversely, biosynthesis of amino acids was enriched in eight modules, indicating that degradation might take place in some conditions, e.g. in DM where the only carbon source had been depleted at the time of sampling, while biosynthesis might take place in the CM where other nutrients from the yeast extract might be consumed. Secondary metabolite biosynthesis was enriched in seven modules and these modules were also enriched for either biosynthesis or degradation of amino acids.

Figure 19. Overview of enriched pathways in coexpression modules. (A) Pathways (rows) enriched among orthologous groups in each coexpression module (columns). (B) Pathway (rows) enriched among genes connected to secondary metabolite backbone genes (columns) in the coexpression modules. Pathways were based on the third level of the MetaCyc annotation, while sub-pathways indicated with arrows, refers to the fourth level of pathway annotation in MetaCyc.

Strong link between primary and secondary metabolism

A total of 33 PKS and NRPS genes were included in the coexpression analysis and all of these were present in at least one coexpression module. To specifically investigate which genes that correlated with secondary metabolism, we conducted an enrichment analysis based on the genes that were directly connected to PKS or NRPS genes in the modules. The genes were divided into either positive or negative correlations based on the PCC. For 20 of the PKS and NRPS genes, at least one pathway was enriched among the correlated genes. The PKS and NRPS genes could be divided into three groups based on the correlated pathways (**Figure 19B**).

Group I correlated mainly with genes involved in degradation pathways, such as amino acid degradation, in particular tyrosine, as well as fatty acid degradation through β-oxidation. In addition, degradation of propanoate, which is a toxic by-product of valine degradation, was correlated via the 2-methylcitrate cycle (2MCC). The only biosynthetic pathways enriched in this group was for secondary metabolites, while a negative correlation to the biosynthesis of amino acids and purine was seen. Group I might constitute genes that are expressed during nutrient limitation, were degradation pathways are active to generate energy and precursor metabolites.

Group II contained NRPS genes that showed a reciprocal correlation pattern compared to group I. Thus, the NRPS genes from group II were negatively correlated to the degradation of amino acids, fatty acids and propanoate. On the other hand, biosynthesis of amino acids and purines were positively correlated to this group. The negative correlation to degradation suggests that this part of secondary metabolism is active during nutrient excess or growth conditions, and thus represent the opposite nutritional requirements compared to group I backbone genes.

For group III, there was only sporadic information of which specific parts of metabolism the PKS and NRPS genes were correlated to. For two NRPSs of this group, a negative correlation to secondary metabolite biosynthesis was observed. This group could constitute secondary metabolite biosynthesis pathways which expression is highly specialized, and dependent on specific environmental stimuli.

Metabolic regulation is tailored for production of secondary metabolites

In our transcriptome analysis, we have identified conserved processes that defines the transcriptional landscape of *Penicillium* during secondary metabolite inducing and nutrient limited conditions. We found that the main pathway that were correlated to the expression of PKS and NRPS genes of group I and II, involved the same pathways that were enriched among differentially expressed gene sets, and among the enriched pathways in the coexpression modules. The pathways that were defining for metabolism of the species, included amino acid degradation in particular of BCAAs and tyrosine, fatty acid degradation through β-oxidation and 2MCC for propanoate degradation (**Figure 16B** and **Figure 19A,B**). The metabolic context of these pathways proved mainly to revolve around the mitochondrial and peroxisomal acetyl-CoA pools, and might constitute the main routes for precursor supply for the secondary metabolism (**Figure 20**).

Precursor formation for secondary metabolite production in fungi has only been studied to a limited extent. Acetyl-CoA generation through β-oxidation of fatty acids has been shown to play an important role in PK biosynthesis (Maggio-Hall et al. 2005). By

disrupting both mitochondrial and peroxisomal β-oxidation individually, the authors observed reduced levels of the PK sterigmatocystin, in *A. nidulans* when grown *in planta*. The mitochondrial β-oxidation and the degradation of valine and isoleucine are further known to be closely linked as they share an enoyl-CoA hydratase, *ECHA*, and an acyl-CoA dehydrogenase, *SCDA* (Maggio-Hall and Keller 2004; Maggio-Hall et al. 2008). Metabolomics profiling of *A. parasiticus* suggested that aside from and β-oxidation, also BCAA degradation contributes towards precursor formation of the PK aflatoxin, and these pathways proved further to be regulated by the global transcriptional regulator of secondary metabolism VeA, (Roze et al. 2010).

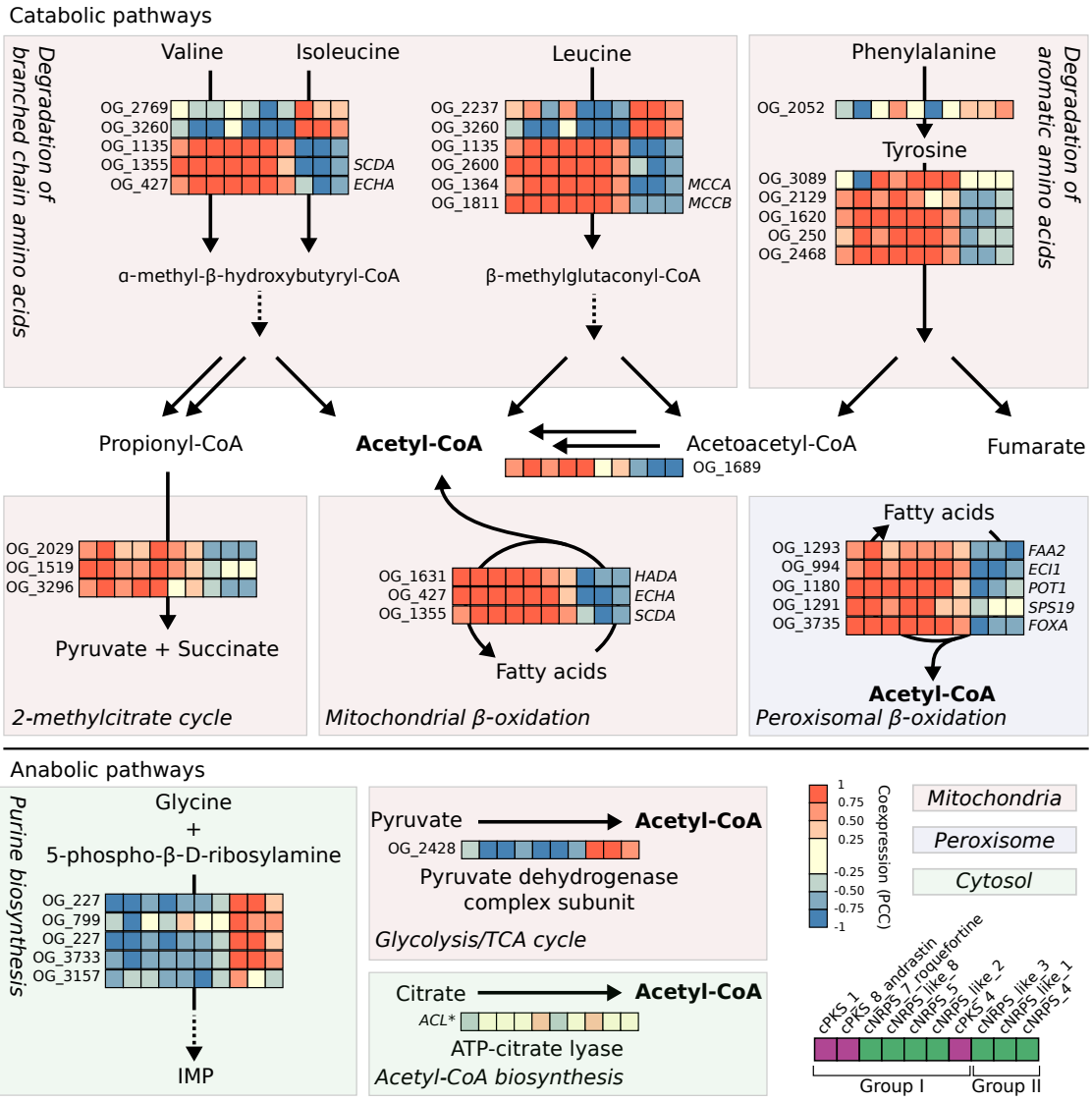


Figure 20. Metabolic pathways correlated with secondary metabolite backbone genes. Heatmaps represent backbone genes in the columns (as specified in the figure key) and orthologous groups (OG) responsible for the catalysis of the reactions in the pathways as rows. The figure key indicates PKSs in purple and NRPSs in green. Coexpression of genes marked with an asterisk (*) was only calculated for five of the species.

Our systematic analysis, combined with previous experimental investigations, suggests a metabolic signature of *Penicillium* during secondary metabolite biosynthetic conditions (**Figure 20**). Since our results are conserved across six diverse *Penicillium* species and

agrees with experimental work conducted in *Aspergillus* species, it is likely to believe that this metabolic signature is conserved across secondary metabolite producing filamentous fungi in general.

Precursors for PKs and NRPs

One important question arises from our analysis: Why does NRPS genes correlate with acetyl-CoA yielding pathways, when NRPs consists of amino acid building blocks and not acetyl-CoA? Global regulators such as the velvet complex (VeA) in *Aspergillus* and *Penicillium* controls the expression of several secondary metabolite pathways at the same time (Becker et al. 2016; Wang et al. 2017). The correlations of NRPS genes with acetyl-CoA generating pathways, might thus be an indirect correlation, based on the common transcriptional activation through global regulators. One could, however, further speculate that the acetyl-CoA production could be important in the biosynthesis of NRPs as well. The by-products from the degradation of BCAAs: acetyl-CoA, glutamate and NADH, constitutes a favourable starting point for the biosynthesis of many amino acids, as it provides the carbon, nitrogen, and reducing power needed. The BCAA degradation could thus be used to supply building blocks for biosynthesis of other amino acids needed for NRP biosynthesis. On the other hand, we did not observe any specific amino acid biosynthesis pathways that correlated with NRPS genes. Another explanation could be that amino acid precursors are generated from degradation of proteins, as it is reasonable to assume that proteins are degraded upon nutrient limitation similarly as observed for fatty acids and amino acids. Possibly, the protein degradation might be the real driver for the precursor supply of secondary metabolism as it would generate amino acids for NRP biosynthesis and BCAAs for acetyl-CoA generation for PK biosynthesis. This is, however, speculative and would require further investigations.

Implications for development of cell factories

Our analysis shows that not only is secondary metabolism correlated to acetyl-CoA generation, but metabolism as a whole is strongly defined by these processes during nutrient limitation. The fact that some of the defining metabolic processes we observed were correlated to secondary metabolism, suggest that filamentous fungi tailor their metabolism to meet the demands for secondary metabolite biosynthesis. This insight suggests that native fungal producers of secondary metabolites have their metabolism optimized for biosynthesis of specific secondary metabolites and their precursors. Based on this, it might be possible to identify wild type fungal species that are naturally optimized for production of specific secondary metabolites, and optimization of such natural producers might thus constitute a promising path towards generation of high yielding industrial cell factories. Further, our results can aid in designing metabolic engineering strategies to optimize production in native secondary metabolite producers e.g. by overexpression of precursor generating pathways.

Paper VI: Metabolic modeling of *Penicillium*

Genome-scale metabolic modeling offer a compelling approach to gain insights into the topology and flux distributions of metabolic networks. Although GEMs have proved powerful in predicting microbial phenotypes (Edwards et al. 2001) and aiding in design of metabolic engineering strategies (Asadollahi et al. 2009), few examples exists where fungal secondary metabolite production has been investigated in the context of GEMs (Nielsen and Nielsen 2017). This lack, has partly been due to a limited understanding of secondary metabolite biosynthesis pathways. With the availability of genome sequences for an increasingly large number of organisms and the improvement of databases containing information on secondary metabolite biosynthesis, the study of secondary metabolism in GEMs is becoming increasingly tractable. The aim of **Paper IV**, was to develop a fast and efficient way of assessing the biosynthetic capabilities of non-model organisms, with limited experimental characterization. We tested a model reconstruction framework on the 24 genome-sequenced *Penicillium* species analyzed in **Paper III**.

GEM reconstruction and metabolic versatility of *Penicillium*

A semi-automatic reconstruction process of GEMs was developed and applied to the 24 species analyzed in **Paper III**. The reconstruction was based on a previously published GEM of *P. rubens* Wisconsin54-1255, as template model (Agren et al. 2013) as well as on the MetaCyc database (Caspi et al. 2014). The advantage of utilizing both resources was the complementarity of the data: The *P. rubens* GEM has been manually curated and contain an accurate description of aspects of metabolism that are not well covered in databases, such as lipid metabolism. The MetaCyc database is comprehensive and allows for addition of reactions to the models that are not in *P. rubens*, and that might not have been experimentally characterized in Penicillia. Additionally, the MetaCyc database has recently undergone a thorough manual curation of secondary metabolite biosynthesis pathways, and a large fraction of secondary metabolism could thus be automatically implemented in the models. Part of the aim of this study was to assess how well this recent development of the MetaCyc database automatically could describe secondary metabolism.

Draft metabolic networks were reconstructed based on inferring orthology between protein sequences of the 24 species, the enzymes in the *P. rubens* GEM and the MetaCyc database, and adding the corresponding reactions to the metabolic networks. A biomass function was added to the individual GEMs by adapting the biomass function of the *P. rubens* GEM and adapting it to the encoded nucleotide and amino acid distributions in the individual genomes. This was followed by a gap-filling process ensuring that the models could produce biomass (Prigent et al. 2017). The resulting functional GEMs contained reactions in the range 2,211-2,658 per model and these were associated with 1,276-1,977 genes, corresponding to roughly 16% of the genomes (**Figure 21**).

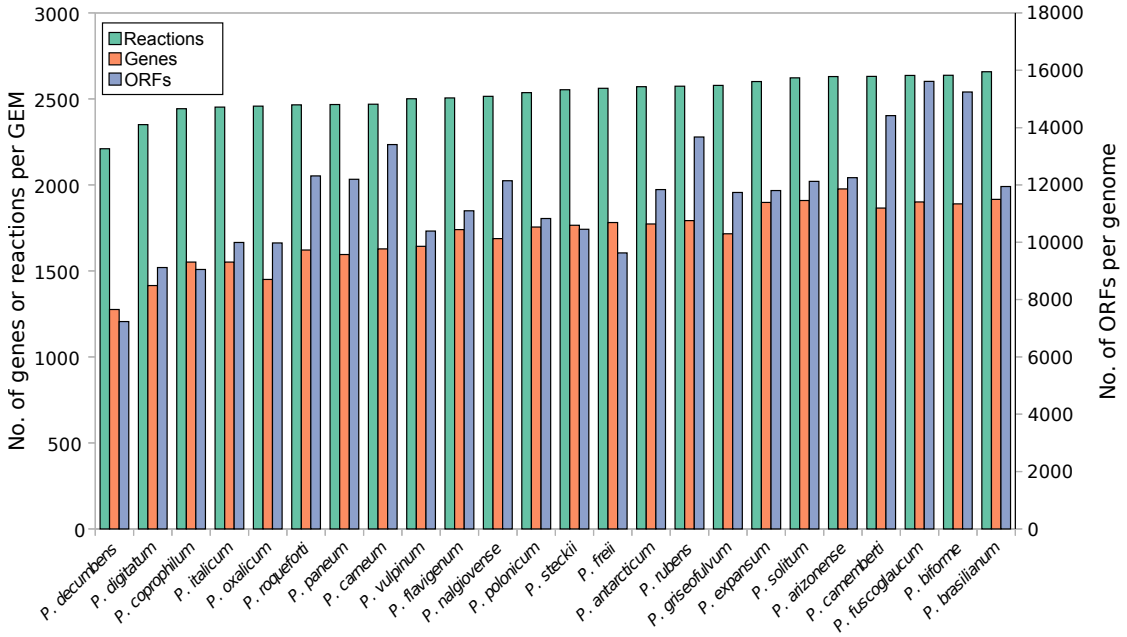


Figure 21. Overview of the size of *Penicillium* GEMs. The GEMs consisted of 2,211-2,658 metabolic reactions and were based on 1,276-1,977 metabolic genes.

Metabolic clustering and functional analysis

The metabolic networks of related organisms are known to be shaped by environmental adaptation (Borenstein et al. 2008) and evolutionary distance (Mazurie et al. 2010). In order to assess how these factors have influenced the metabolism of the *Penicillium* species, we compared a hierarchical clustering of the metabolic networks, to the phylogeny and habitats of the species (**Figure 22**). The clustering of the metabolic networks, based on the presence of reactions, was divided into eight metabolic clades, and these proved in most cases to be in accordance with the phylogenetic clades. However, some noteworthy differences were observed: For example, *P. expansum* is a pome fruit pathogen, and grouped with another pome fruit pathogen from a different phylogenetic clade, *P. solitum*. Here the shared habitats seemed to have resulted in a metabolic similarity between these two species, compared to *P. italicum* and *P. digitatum* that are citrus fruit pathogens, but phylogenetically related to *P. expansum*. By comparing the reactions unique to each of the two groups, some reactions in sphingolipid metabolism proved to be missing in *P. expansum* and *P. solitum*. Another interesting observation was the metabolic grouping of *P. griseofulvum* and *P. flavigenum* that swapped position relative to their phylogenetic groups. We could not correlate this change to any habitat preferences or any major differences in metabolic pathways that differentiated the species in the two clades.

We further performed an enrichment analysis between all metabolic clades, to assess which pathways that were enriched among reactions that differentiated clades, relative to the pan-reactome, i.e. the union of all reactions in the 24 GEMs (hypergeometric test, $p<0.01$). We found some notable differences in metabolic clade 3 that proved to contain unique reactions for degradation of chlorobenzenes, while metabolic clade 3 and 6 where depleted for reactions involved in creatinine degradation. Secondary metabolism, however, proved to be the main driver of the metabolic grouping and differences

observed, in agreement with the findings in **Paper III**, that showed that secondary metabolism constitute the greatest variation within the genomes.

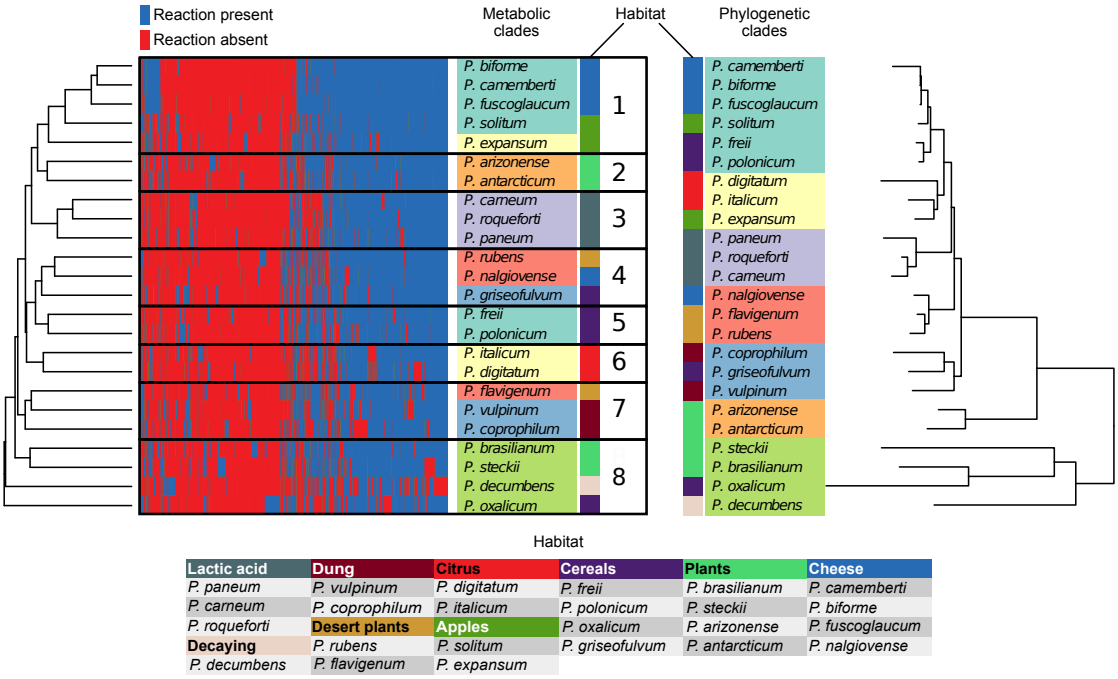


Figure 22. Clustering of *Penicillium* species based on metabolic reactions and phylogeny. The presence or absence of metabolic reactions were used to cluster the species into eight different metabolic clades. The metabolic clades were compared to phylogenetic clades and the habitats of the species.

Secondary metabolite pathways in GEMs

One of the goals of this study was to evaluate the accuracy of the annotation of the secondary metabolism obtained by the automatic reconstruction processes. A total of 33 different pathways associated to secondary metabolism according to MetaCyc were present in the 24 GEMs (**Figure 23**). Among the predicted pathways, several proved to accurately agree with known species:metabolite associations (Frisvad et al. 2004; Nielsen et al. 2017). These included conserved fungal pathways, such as geranylgeranyl diphosphate, mevalonate and ergotamine biosynthesis, as well as some pathways that are more specific, such as patulin, penicillin, tryptoquialanine, griseofulvin and aurofusarin biosynthesis. Other pathways were clearly false positives, and over-predicted in abundance such as stipitatate biosynthesis which was predicted to be present in 14 species, but only production of the related compound puberulonic acid is known to be taking place in some of the species (*P. freii* and *P. polonicum*) (Nielsen et al. 2017). Taken together, the automatic annotation of secondary metabolism, can be used as a good starting point for identifying many secondary metabolite pathways, but it does require subsequent curation to obtain a reliable description of the secondary metabolism.

FBA simulations were conducted for all *Penicillium* GEMs for the production of the PK griseofulvin and the NRP penicillin. Biosynthesis pathways were added to all models and maximized based on glucose as the sole carbon source and biomass production being fixed. The maximum theoretical yields proved to be very similar among all models: 0.43 mole penicillin/mole glucose, and 0.35 mole griseofulvin/mole glucose.

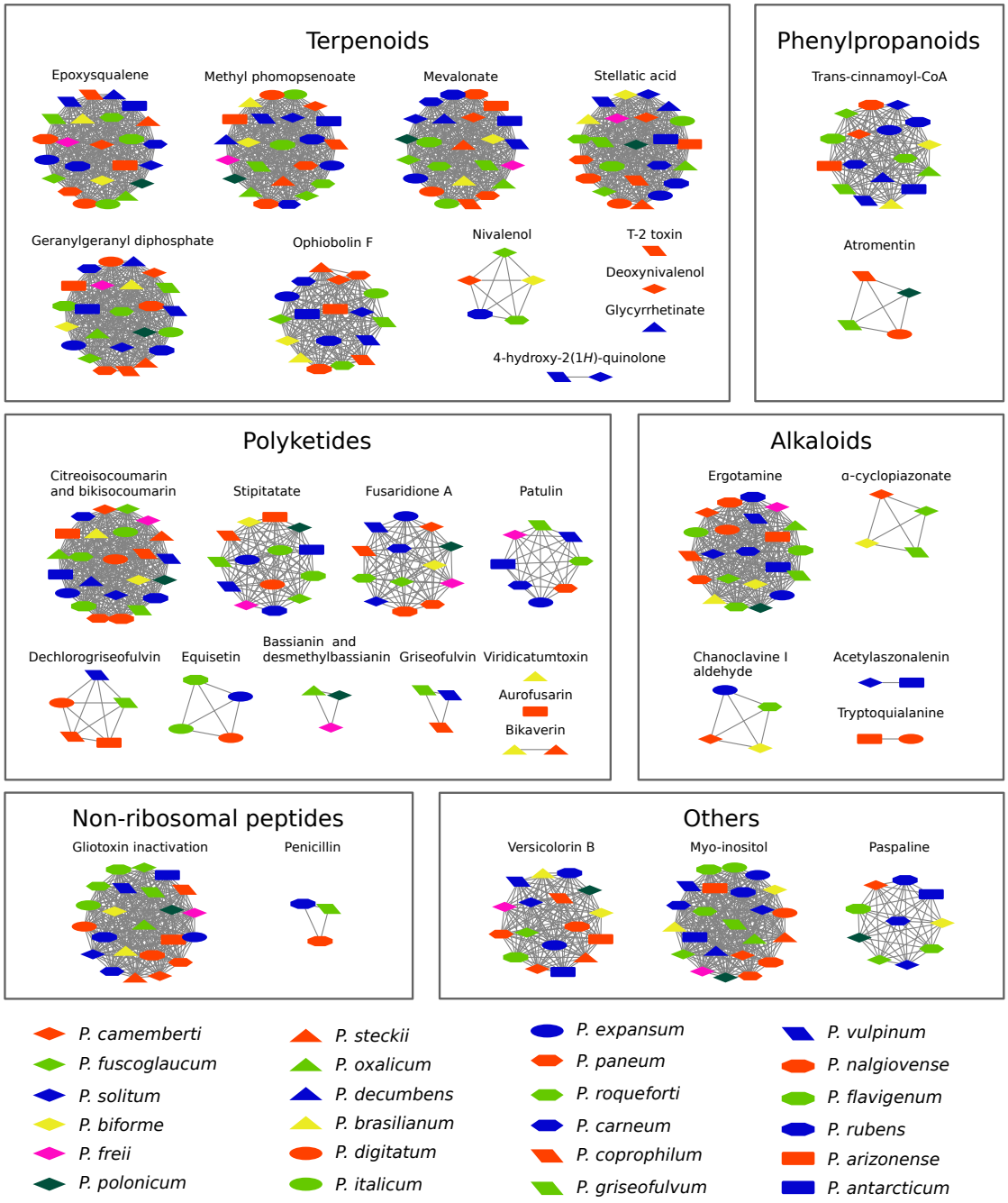


Figure 23. Secondary metabolite biosynthesis predicted in the *Penicillium* GEMs. Each node represents a species and the corresponding compounds is given on top of each clusters. All secondary metabolites are grouped into boxes of chemical classes.

It is interesting to speculate how one could improve these predictions, in order to identify species that might possess a greater potential for secondary metabolite production. According to our network analysis of metabolism, central carbon metabolism is very conserved among the species, and since precursor formation for secondary metabolism takes place in the central carbon metabolism, this might be the reason for the observed similarities in secondary metabolite production. Regulations that activate/repress different pathways in metabolism, might be a way fungi attain metabolic differences, while the full metabolic potential is conserved. Determining which pathways are active could be identified by integrating omics data in the network analysis (Machado and Herrgård 2014). Another factor that might be important in particular in secondary

metabolism is the costs of protein synthesis relative to the activity of the enzymes. PKSs and NRPSs are known to be very large enzymes with a low catalytic activity (Thykaer and Nielsen 2003), and these parameters indicate that enzyme biosynthesis could be a major energy expenditure in secondary metabolite production. In yeast, the development of an enzyme constrained metabolic model, has shown to increase the accuracy of predictions, and accurately describe experimental data (Sánchez et al. 2017). Thus, the implementation of such kinetic enzyme parameters in GEMs of species that synthesize secondary metabolites might be a way of improving phenotype predictions related to secondary metabolism.

In this part, I have shown that *Penicillium* biodiversity shows potential for industrial exploitations as secondary metabolite producing cell factories. A physiological characterization suggests that uncharacterized *Penicillium* species can be cultivated and perform well in submerged bioreactor fermentations in terms of growth and production of secondary metabolites. Further, these wild type species holds an intrinsic metabolic potential for secondary metabolite production and further optimization. The availability of GEMs of these species provides the basis for optimization studies of metabolism and enables further studies of *Penicillium* metabolism.

Conclusions and perspectives

This thesis constitutes a thorough investigation of the capabilities of filamentous fungi for production of secondary metabolites. In particular, the fungal genus of *Penicillium* has been the center of attention in this endeavour to increase the understanding of genetically encoded biosynthesis pathways, the physiology and metabolism of secondary metabolite production.

In part I of the thesis, two studies showcased methods for connecting BGCs to secondary metabolites as reviewed in **Paper I**. In **Paper II**, the overall aim was to describe a new fungal species, *Penicillium arizonense*. We were interested in not only providing the basis for future identification of the species, but also enabling further biotechnological exploitation. We identified a genetic potential for production of a large number of secondary metabolites as well as degradative enzymes. Seven different families of secondary metabolites were detected in culture extracts of the species, and many of these had bioactive properties of pharmaceutical relevance. For six of these compounds, we were able to link them to BGCs identified in the genome. Our approach, constituted a low-throughput, but accurate evaluation of secondary metabolite biosynthesis in a single species. In **Paper III**, we developed a high-throughput automated pipeline to group and annotate BGCs in 24 *Penicillium* genomes. The automatic BGC annotation was followed by an evaluation of the accuracy of the predictions and the majority (89 of 127) of the species:metabolite associations could be validated based on literature or chemical analysis conducted in the study. Our analysis demonstrated that only 16% of PKS and NRPS BGCs in 24 *Penicillium* genomes could be assigned to a pathway, thus highlighting a major untapped biosynthetic potential.

In total, we found 1,317 BGCs in the 24 genomes and this high number makes it evident that the characterization of new BGC cannot keep pace with genome mining approaches identifying orphan BGCs. Consequently, a major challenge for future secondary metabolite research is the prioritization of such BGCs in the hunt for new pharmaceuticals. In our study, we decided to focus on yanuthones, since the presence of an extra gene in the *Penicillium* version of the BGC suggested that the pathway might be extended compared to the characterized pathway in *A. niger* (Holm et al. 2014). A recent study identified an extended austinol pathway in *A. calidoustus* and showed that new austinol compounds from the pathway had more specific insecticidal activity compared to austinols known from *A. nidulans* (Valiante et al. 2017). We identified a novel compound from the yanuthone pathway, but further studies are required to evaluate if this compound has increased specificity or potency compared to the antifungal properties of yanuthones from the *A. niger* pathway.

This method of identifying extended pathways, doesn't allow for exploration of unknown pathways and thus does not provide a fully satisfying solution to the problem of prioritizing BGCs. Self-resistance genes, however, could potentially be used for predicting the activity of the end-product of a secondary metabolite pathway. The benefit of self-resistance genes is that they can easily be identified in the genome and it can be evaluated if they map to a BGC. Features such as known resistance genes, essential core

genes and horizontal gene transfer events, are used in the ARTS server mentioned previously, to identify potential self-resistance genes and prioritize BGCs that could encode antibiotic pathways, and this is a major step towards more targeted BGC mining for antibiotic discovery (Alanjary et al. 2017). The ARTS server is however, specialized towards bacterial gene clusters, and since most secondary metabolite research is based on bacteria, efforts should be made to adapt such tools to fungal specific parameters as well.

In part II, the focus was more applied and aimed at elucidating fungal physiology in relation to secondary metabolite production. Bioreactor batch fermentations were conducted in **Paper IV** and revealed that all ten genome-sequenced *Penicillium* species grew in a reproducible and exponential fashion in rich medium, and produced secondary metabolites. Considering that submerged conditions is not a natural state for most *Penicillium* species, it was encouraging to find that all species showed beneficial growth characteristics that suggest the use of native secondary metabolite producers as potential cell factories. The majority of secondary metabolite studies in fungi is conducted on solid media, and we found that fewer secondary metabolites were produced under submerged conditions. This reduced number of produced secondary metabolites might be beneficial for production, since competition from other secondary metabolite pathways is reduced. Thus, species that grow fast and produce few secondary metabolites in submerged conditions might constitute a good starting point to establish a production process.

Further analysis of the cultivated species using comparative transcriptomics in **Paper V**, showed that metabolism of filamentous fungi is tailored towards secondary metabolite production, and this might be an important realization in the development of new cell factories. The successful optimization of penicillin production by random mutagenesis of *P. rubens* (van den Berg et al. 2008), might be directly attributable to the fact that metabolism of filamentous fungi is natively wired for secondary metabolite production, and thus can be further improved with only few genetic modifications. Conversely, over-production of secondary metabolites in a heterologous host such as *S. cerevisiae*, that is not natively producing secondary metabolites, might prove difficult as it will require fundamental rewiring of metabolism to achieve the same titres as native producers. Further, there is a striking lack of examples of heterologous expression of secondary metabolite pathways in yeast, which suggests that some fundamental features for secondary metabolite production is missing, e.g. compartmentalization of biosynthesis steps (Kistler et al. 2015). An alternative could be heterologous production in model secondary metabolite producers such has *Aspergillus* species (Anyaogu and Mortensen 2015) or *P. rubens*, that has already been improved for production of NRPs (penicillin) (van den Berg et al. 2008).

A transcriptional coexpression analysis further identified a metabolic signature for secondary metabolite biosynthesis, in particular acetyl-CoA precursor supply for PKs during nutrient limitation. Our findings were strongly supported by previous studies on *Aspergillus* species suggesting mitochondrial acetyl-CoA through BCAA and fatty acid degradation to constitute important pathways for secondary metabolite precursors (Roze et al. 2010). We extended these precursor-generating pathways to include tyrosine degradation as well. The conservation across *Penicillium* and *Aspergillus* species, suggests that these pathways are generally responsible for secondary metabolite precursor formation in filamentous fungi. Utilizing this knowledge to develop metabolic engineering strategies, e.g. by overexpressing precursor pathways, could be a promising strategy to increase the production of secondary metabolites in native producers.

In **Paper IV**, we focused specifically on metabolism of the *Penicillium* genus by reconstructing GEMs for the 24 species analyzed in **Paper III.** The aim was to evaluate if metabolism could be accurately reconstructed for genome sequenced organisms with minor experimental characterization. We found that the metabolic networks of the 24 species were very similar, but showed variations in secondary metabolite biosynthesis pathways. The similarities in central metabolism of the species also manifested in very similar maximum theoretical yields of penicillin and griseofulvin based on FBA simulations of the models. Future improvements of model accuracy, e.g. by implementing kinetic parameters, might increase the predictive power of the GEMs, and constitute a promising way of screening which species has the better intrinsic capabilities for production of specific secondary metabolites.

In the introduction of this thesis, I stated that the aim was to contribute to the development of novel antibiotics. Although no new antibiotics were experimentally validated in the work presented here, I would argue that most likely a number of new antibiotics have been found. I highlighted the diversity in secondary metabolite biosynthesis pathways in the *Penicillium* genus, and defined which species contain which pathways. Future developments of prioritization strategies among the detected BGC will push this work forward, as discussed above. I further suggest that despite developments in establishing platform cell factories such as *S. cerevisiae* for production of chemicals, native secondary metabolite producers might still represent efficient platforms that are naturally geared for production. I hope that the work presented in this thesis has contributed to the understanding of fungal secondary metabolite production, and will facilitate future development of fungal derived antibiotics.

Acknowledgements

I would like to thank Professor Jens Nielsen for supervision and guidance throughout my PhD project. I am grateful for having been part of your research group, and working with you has always been extremely motivating and inspiring. My two co-supervisors Dr. Sylvain Prigent and Dr. Boyang Ji are acknowledged for always being open for discussions and for their many contributions to my project.

The QuantFung consortium constituted an excellent platform to discuss the biology of fungal secondary metabolite producers. It was a pleasure to interact with many of the leading research groups within the field.

I had a very fruitful collaboration with DTU, and I acknowledge Associate Professor Mhairi Workman and Professor Uffe Mortensen for enabling my exchange. I had inspiring collaborations with Dr. Sietske Grijseels, Associate Professor Kristian Fog Nielsen, Associate Professor Rasmus Frandsen and Professor Jens Christian Frisvad.

The Systems and Synthetic Biology division consists of a dynamic, multicultural and yet uniform group of people, that all share a passion for science and these people constituted a great work place where it was always exciting to go to work. I met many personalities here that I won't forget.

Julie thanks for travelling with me to Sweden and for always getting the best out of me.

References

Agren R, Liu L, Shoaie S, et al (2013) The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum. PLoS Comput Biol 9:e1002980. doi: 10.1371/journal.pcbi.1002980

Alanjary M, Kronmiller B, Adamek M, et al (2017) The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. Nucleic Acids Res 45:W42–W48. doi: 10.1093/nar/gkx360

Aminov RI (2010) A Brief History of the Antibiotic Era: Lessons Learned and Challenges for the Future. Front Microbiol 1:134. doi: 10.3389/fmicb.2010.00134

Andersen MR, Nielsen JB, Klitgaard A, et al (2013) Accurate prediction of secondary metabolite gene clusters in filamentous fungi. Proc Natl Acad Sci 110:E99–E107. doi: 10.1073/pnas.1205532110

Anyaogu DC, Mortensen UH (2015) Heterologous production of fungal secondary metabolites in Aspergilli. Front Microbiol 6:77. doi: 10.3389/fmicb.2015.00077

Artigot MP, Loiseau N, Laffitte J, et al (2009) Molecular cloning and functional characterization of two CYP619 cytochrome P450s involved in biosynthesis of Patulin in Aspergillus clavatus. Microbiology 155:1738–1747. doi: 10.1099/mic.0.024836-0

Asadollahi MA, Maury J, Patil KR, et al (2009) Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering. Metab Eng 11:328–334. doi: 10.1016/j.ymben.2009.07.001

Ashburner M, Ball CA, Blake JA, et al (2000) Gene Ontology: tool for the unification of biology. Nat Genet 2000 251 25:25–29. doi: doi:10.1038/75556

Awan AR, Blount BA, Bell DJ, et al (2017) Biosynthesis of the antibiotic nonribosomal peptide penicillin in baker's yeast. Nat Commun 8:15202. doi: 10.1038/ncomms15202

Bakri Y, Jacques P, Thonart P (2003) Xylanase production by Penicillium canescens 10-10c in solid-state fermentation. Appl Biochem Biotechnol 105–108:737–748. doi: 10.1385/ABAB:108:1-3:737

Ballester A, Marcet-houben M, Levin E, et al (2015) Genome, Transcriptome, and Functional Analyses of Penicillium expansum Provide New Insights Into Secondary Metabolism and Pathogenicity. Mol Plant-Microbe Interact 28:232–248. doi: 10.1094/MPMI-09-14-0261-FI

Banani H, Marcet-Houben M, Ballester A-R, et al (2016) Genome sequencing and secondary metabolism of the postharvest pathogen Penicillium griseofulvum. BMC Genomics 17:19. doi: 10.1186/s12864-015-2347-x

Barrios-González J, Miranda RU (2010) Biotechnological production and applications of statins. Appl Microbiol Biotechnol 85:869–883. doi: 10.1007/s00253-009-2239-6

Bayram Ö, Krappmann S, Ni M, et al (2008) VelB/VeA/LaeA Complex Coordinates Light Signal with Fungal Development and Secondary Metabolism. Science 320:1504–1506.

Becker K, Ziemons S, Lentz K, et al (2016) Genome-Wide Chromatin Immunoprecipitation Sequencing Analysis of the Penicillium chrysogenum Velvet Protein PcVelA Identifies Methyltransferase PcLlmA as a Novel Downstream Regulator of Fungal Development. Mol Biol Physiol 1:e00149-16. doi: 10.1128/mSphere.00149-16

Bérdy J (2005) Bioactive microbial metabolites. J Antibiot 58:1–26. doi: 10.1038/ja.2005.1

Bok JW, Noordermeer D, Kale SP, Keller NP (2006) Secondary metabolic gene cluster silencing in Aspergillus nidulans. Mol Microbiol 61:1636–1645. doi: 10.1111/j.1365-2958.2006.05330.x

Borenstein E, Kupiec M, Feldman MW, Ruppin E (2008) Large-scale reconstruction and phylogenetic analysis of metabolic environments. Proc Natl Acad Sci U S A 105:14482–7. doi: 10.1073/pnas.0806162105

Brakhage AA (1998) Molecular regulation of beta-lactam biosynthesis in filamentous fungi. Microbiol Mol Biol Rev 62:547–85.

Brakhage AA (2013) Regulation of fungal secondary metabolism. Nat Rev Microbiol 11:21–32. doi: 10.1038/nrmicro2916

Brandl J, Andersen MR (2017) Aspergilli: Models for systems biology in filamentous fungi. Curr Opin Syst Biol 6:67–73. doi: 10.1016/j.coisb.2017.09.005

Butler MJ, Day AW (1998) Fungal melanins: a review. Can J Microbiol 44:1115–1136. doi: 10.1139/w98-119

Büttel Z, Díaz R, Dirnberger B, et al (2015) Unlocking the potential of fungi: the QuantFung project. Fungal Biol Biotechnol 2:6. doi: 10.1186/s40694-015-0016-0

Cacho RA, Tang Y, Chooi YH (2015) Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi. Front Microbiol 6:1–16. doi: 10.3389/fmicb.2014.00774

Caspi R, Altman T, Billington R, et al (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res 42:D459–D471. doi: 10.1093/nar/gkt1103

Chain E, Florey HW, Gardner AD, et al (1940) Penicillin as a chemoterapeutic agent. Lancet 236:226–228. doi: 10.1016/S0140-6736(01)08728-1

Cheeseman K, Ropars J, Renault P, et al (2014) Multiple recent horizontal transfers of a large genomic region in cheese making fungi. Nat Commun 5:2876. doi: 10.1038/ncomms3876

Chooi Y-H, Cacho R, Tang Y (2010) Identification of the Viridicatumtoxin and Griseofulvin Gene Clusters from Penicillium aethiopicum. Chem Biol 17:483–494. doi: 10.1016/j.chembiol.2010.03.015

Christensen T, Woeldike H, Boel E, et al (1988) High level expression of recombinant genes in Aspergillus oryzae. Nat Biotechnol 6:1419–1422. doi: 10.1038/nbt1288-1419

Cimermancic P, Medema MH, Claesen J, et al (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158:412–421. doi: 10.1016/j.cell.2014.06.034

Conway KR, Boddy CN (2013) ClusterMine360: A database of microbial PKS/NRPS biosynthesis. Nucleic Acids Res 41:402–407. doi: 10.1093/nar/gks993

Crous PW, Verkley GJM, Groenewald JZ, Samson RA (2009) Fungal Biodiversity, 1st edn. CBS-KNAW Fungal Biodiversity Centre, Utrecht

Currie JN (1917) The citric acid fermentation of Aspergillus niger. J Biol Chem 31:15.

DaSilva, J. E (2012) The Colours of Biotechnology: Science, Development and Humankind. Electron J Biotechnol. doi: 10.2225/vol7-issue3-fulltext-

Davies J, Davies D (2010) Origins and Evolution of Antibiotic Resistance. Microbiol Mol Biol Rev 74:417–433. doi: 10.1128/MMBR.00016-10

de Jesus AE, Horak RM, Steyn PS, Vleggaar R (1983) Biosynthesis of austalide D, a meroterpenoid mycotoxin from Aspergillus ustus. J Chem Soc Chem Commun 716–718. doi: 10.1039/c39830000716

de Queiroz A, Gatesy J (2007) The supermatrix approach to systematics. Trends Ecol Evol 22:34–41. doi: 10.1016/j.tree.2006.10.002

de Vries RP, Riley R, Wiebenga A, et al (2017) Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus Aspergillus. Genome Biol 18:28. doi: 10.1186/s13059-017-1151-0

Dejong CA, Chen GM, Li H, et al (2016) Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. Nat Chem Biol 12:1007–1014. doi: 10.1038/nchembio.2188

Doroghazi JR, Albright JC, Goering AW, et al (2014) A roadmap for natural product discovery based on large-scale genomics and metabolomics. Nat Chem Biol 10:963–8. doi: 10.1038/nchembio.1659

Edwards JS, Ibarra RU, Palsson BØ (2001) In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat Biotechnol 19:125–130. doi: 10.1038/84379

Finkelstein E, Amichai B, Grunwald MH (1996) Griseofulvin and its uses. Int J Antimicrob Agents 6:189–194. doi: 10.1016/0924-8579(95)00037-2

Fischer S, Brunk BP, Chen F, et al (2011) Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster Proteomes Into New Ortholog Groups. Curr Protoc Bioinforma 35:6.12:6.12.1–6.12.19. doi: 10.1002/0471250953.bi0612s35

Fleming A (1929) On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. Br Jounal Exp Pathol 10:226–236.

Frandsen RJN, Schütt C, Lund BW, et al (2011) Two novel classes of enzymes are required for the biosynthesis of aurofusarin in Fusarium graminearum. J Biol Chem 286:10419–28. doi: 10.1074/jbc.M110.179853

Frisvad JC (2012) Media and Growth Conditions for Induction of Secondary Metabolite Production. In: keller N, Turner G (eds) Fungal Secondary Metabolism. Humana Press, Totowa, NJ, pp 47–58

Frisvad JC, Smedsgaard J, Larsen TO, Samson RA (2004) Mycotoxins, drugs and other extrolites produced by species in Penicillium subgenus Penicillium. Stud Mycol 49:201–241.

Galagan JE, Calvo SE, Borkovich KA, et al (2003) The genome sequence of the filamentous fungus Neurospora crassa. Nature 422:859–868. doi: 10.1038/nature01554

Galagan JE, Calvo SE, Cuomo C, et al (2005) Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae. Nature 438:1105–15. doi: 10.1038/nature04341

Gao X, Chooi Y-H, Ames BD, et al (2011) Fungal indole alkaloid biosynthesis: genetic and biochemical investigation of the tryptoquialanine pathway in Penicillium aethiopicum. J Am Chem Soc 133:2729–41. doi: 10.1021/ja1101085

Gomez-Escribano JP, Bibb MJ (2011) Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters. Microb Biotechnol 4:207–215. doi: 10.1111/j.1751-7915.2010.00219.x

Gomez R, Schnabel I, Garrido J (1988) Pellet growth and citric add yield of Aspergillus niger 110. Enzyme Microb Technol 10:188–191.

Gressler M, Meyer F, Heine D, et al (2015) Phytotoxin production in Aspergillus terreus is regulated by independent environmental signals. Elife 4:1–29. doi: 10.7554/eLife.07861

Grigoriev I V, Nikitin R, Haridas S, et al (2014) MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res 42:D699-704. doi: 10.1093/nar/gkt1183

Grijseels S, Nielsen JC, Randelovic M, et al (2016) Penicillium arizonense, a new, genome sequenced fungal species, reveals a high chemical diversity in secreted metabolites. Sci Rep 6:35112. doi: 10.1038/srep35112

Gusakov A V (2011) Alternatives to Trichoderma reesei in biofuel production. Trends Biotechnol 29:419–425. doi: 10.1016/j.tibtech.2011.04.004

Hadjithomas M, Chen IA, Chu K, et al (2015) IMG-ABC: A Knowledge Base To Fuel Discovery of Biosynthetic Gene Clusters and Novel Secondary Metabolites. MBio 6:e00932-15. doi: 10.1128/mBio.00932-15

Hillman ET, Readnour LR, Solomon K V (2017) Exploiting the natural product potential of fungi with integrated -omics and synthetic biology approaches. Curr Opin Syst Biol 5:50–56. doi: 10.1016/j.coisb.2017.07.010

Hoffmeister D, Keller NP (2007) Natural products of filamentous fungi: enzymes, genes, and their regulation. Nat Prod Rep 24:393–416. doi: 10.1039/B603084J

Holm DK, Petersen LM, Klitgaard A, et al (2014) Molecular and chemical characterization of the biosynthesis of the 6-MSA-derived meroterpenoid yanuthone D in Aspergillus niger. Chem Biol 21:519–529. doi: 10.1016/j.chembiol.2014.01.013

Houbraken J, Wang L, Lee HB, Frisvad JC (2015) New sections in Penicillium containing novel species producing patulin, pyripyropens or other bioactive compounds. Persoonia 36:299–314. doi: 10.3767/003158516X692040

Itoh T, Tokunaga K, Matsuda Y, et al (2010) Reconstitution of a fungal meroterpenoid biosynthesis reveals the involvement of a novel family of terpene cyclases. Nat Chem 2:858–64. doi: 10.1038/nchem.764

Jakociunas T, Jensen MK, Keasling JD (2016) CRISPR/Cas9 advances engineering of microbial cell factories. Metab Eng 34:44–59. doi: 10.1016/J.YMBEN.2015.12.003

Kanehisa M, Sato Y, Kawashima M, et al (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457–D462. doi: 10.1093/nar/gkv1070

Keller NP, Turner G, Bennett JW (2005) Fungal secondary metabolism - from biochemistry to genomics. Nat Rev Microbiol 3:937–947. doi: 10.1038/nrmicro1286

Khaldi N, Seifuddin FT, Turner G, et al (2010) SMURF: Genomic mapping of fungal secondary metabolite clusters. Fungal Genet Biol 47:736–741. doi: 10.1016/j.fgb.2010.06.003

Kinoshita S, Udaka S, Shimino M (1957) Studies on the amino acid fermentation. Part I. Production of L-glutamic acid by various organisms. J Gen Appl Microbiol 3:193–205.

Kistler HC, Broz K, Yu J-H, Linz JE (2015) Cellular compartmentalization of secondary metabolism. 6:68. doi: 10.3389/fmicb.2015.00068

Kitano H (2002a) Systems Biology: A Brief Overview. Science 295:1662–1664.

Kitano H (2002b) Computational systems biology. Nature 420:206–210.

Komatsu M, Uchiyama T, Omura S, et al (2010) Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. Proc Natl Acad Sci U S A 107:2646–51. doi: 10.1073/pnas.0914833107

Kresge N, Simoni RD, Hill RL (2004) Selman Waksman: The Father of Antibiotics. J Biol Chem 279:e7.

Krijgsheld P, Bleichrodt R, van Veluw GJ, et al (2013) Development in Aspergillus. Stud Mycol 74:1–29. doi: 10.3114/sim0006

Kroken S, Glass NL, Taylor JW, et al (2003) Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. Proc Natl Acad Sci U S A 100:15670–15675. doi: 10.1073/pnas.2532165100

Lawrence JG, Roth JR (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics 143:1843–60.

Li X, Choi HD, Kang JS, et al (2003) New Polyoxygenated Farnesylcyclohexenones, Deacetoxyyanuthone A and Its Hydro Derivative from the Marine-Derived Fungus Penicillium sp. Jounal Nat Prod 66:1499–1500. doi: 10.1021/NP030231U

Li YF, Tsai KJS, Harvey CJB, et al (2016) Comprehensive curation and analysis of fungal biosynthetic gene clusters of published natural products. Fungal Genet Biol 89:18–28. doi: 10.1016/j.fgb.2016.01.012

Lind AL, Wisecaver JH, Lameiras C, et al (2017) Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species. PLoS Biol 15:e2003583. doi: 10.1371/journal.pbio.2003583

Liu G, Qin Y, Li Z, Qu Y (2013a) Improving lignocellulolytic enzyme production with Penicillium: from strain screening to systems biology. Biofuels 4:523–534. doi: 10.4155/bfs.13.38

Liu G, Zhang L, Wei X, et al (2013b) Genomic and Secretomic Analyses Reveal Unique Features of the Lignocellulolytic Enzyme System of Penicillium decumbens. PLoS One 8:e55185. doi: 10.1371/journal.pone.0055185

Love MI, Huber W, Anders S, et al (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. doi: 10.1186/s13059-014-0550-8

Lucht JM (2015) Public Acceptance of Plant Biotechnology and GM Crops. Viruses 7:4254–4281. doi: 10.3390/v7082819

Machado D, Herrgård M (2014) Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism. PLoS Comput Biol 10:e1003580. doi: 10.1371/journal.pcbi.1003580

Machida M, Asai K, Sano M, et al (2005) Genome sequencing and analysis of Aspergillus oryzae. Nature 438:1157–61. doi: 10.1038/nature04300

Maggio-Hall LA, Keller NP (2004) Mitochondrial β-oxidation in Aspergillus nidulans. Mol Microbiol 54:1173–1185. doi: 10.1111/j.1365-2958.2004.04340.x

Maggio-Hall LA, Lyne P, Wolff JA, Keller NP (2008) A single acyl-CoA dehydrogenase is required for catabolism of isoleucine, valine and short-chain fatty acids in Aspergillus nidulans. Fungal Genet Biol 45:180–189. doi: 10.1016/j.fgb.2007.06.004

Maggio-Hall LA, Wilson RA, Keller NP (2005) Fundamental contribution of beta-oxidation to polyketide mycotoxin production in planta. Mol Plant-Microbe Interact 18:783–793. doi: 10.1094/MPMI-18-0783

Marcet-Houben M, Ballester A-R, de la Fuente B, et al (2012) Genome sequence of the necrotrophic fungus Penicillium digitatum, the main postharvest pathogen of citrus. BMC Genomics 13:646. doi: 10.1186/1471-2164-13-646

Martinez D, Berka RM, Henrissat B, et al (2008) Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina). Nat Biotechnol 26:553–560. doi: 10.1038/nbt1403

Maskey RP, Grün-Wollny I, Laatsch H (2005) Sorbicillin Analogues and Related Dimeric Compounds from Penicillium notatum. J Nat Prod 68:865–870. doi: 10.1021/NP040137T

Mazurie A, Bonchev D, Schwikowski B, Buck GA (2010) Evolution of metabolic network organization. BMC Syst Biol 4:59.

McDonagh A, Fedorova ND, Crabtree J, et al (2008) Sub-Telomere Directed Gene Expression during Initiation of Invasive Aspergillosis. PLoS Pathog 4:e1000154. doi: 10.1371/journal.ppat.1000154

McGovern PE, Zhang J, Tang J, et al (2004) Fermented beverages of pre- and proto-historic China. Proc Natl Acad Sci U S A 101:17593–17598. doi: 10.1073/pnas.0407921102

Medema MH, Blin K, Cimermancic P, et al (2011) AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39:339–346. doi: 10.1093/nar/gkr466

Medema MH, Fischbach MA (2015) Computational approaches to natural product discovery. Nat Chem Biol 11:639–648. doi: 10.1038/nchembio.1884

Medema MH, Kottmann R, Yilmaz P, et al (2015) Minimum Information about a Biosynthetic Gene cluster. Nat Chem Biol 11:625–631. doi: 10.1038/nchembio.1890

Medema MH, van Raaphorst R, Takano E, Breitling R (2012) Computational tools for the synthetic design of biochemical pathways. Nat Rev Microbiol 10:191–202. doi: 10.1038/nrmicro2717

Nepusz T, Yu H, Paccanaro A (2012) Detecting overlapping protein complexes in protein-protein interaction networks. Nat Methods 9:471–472. doi: 10.1038/Nmeth.1938

Nielsen J (2001) Metabolic engineering. Appl Microbiol Biotechnol 55:263–283. doi: 10.1007/s002530000511

Nielsen J (2017) Systems Biology of Metabolism. Annu Rev Biochem 86:245–275. doi: 10.1146/annurev-biochem-061516-044757

Nielsen J (2014) Synthetic Biology for Engineering Acetyl Coenzyme A Metabolism in Yeast. MBio 5:e02153-14. doi: 10.1128/mBio.02153-14

Nielsen J, Keasling JD (2016) Engineering Cellular Metabolism. Cell 164:1185–1197. doi: 10.1016/j.cell.2016.02.004

Nielsen JC, Grijseels S, Prigent S, et al (2017) Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in Penicillium species. Nat Microbiol 2:17044. doi: 10.1038/nmicrobiol.2017.44

Nielsen JC, Nielsen J (2017) Development of fungal cell factories for the production of secondary metabolites: linking genomics and metabolism.

Nierman WC, Pain A, Anderson MJ, et al (2005) Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. Nature 438:1151–1156. doi: 10.1038/nature04332

Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis. Nat Biotechnol 28:245–248.

Palmer JM, Keller NP (2010) Secondary metabolism in fungi: does chromosomal location matter? Curr Opin Microbiol 13:431–436. doi: 10.1016/j.mib.2010.04.008

Patil KR, Nielsen J (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proc Natl Acad Sci U S A 102:2685–9. doi: 10.1073/pnas.0406811102

Petersen LM, Holm DK, Knudsen PB, et al (2015) Characterization of four new antifungal yanuthones from Aspergillus niger. J Antibiot 68:201–205. doi: 10.1038/ja.2014.130

Price ND, Reed JL, Palsson BØ (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol 2:886–897. doi: 10.1038/nrmicro1023

Prigent S, Frioux C, Dittami SM, et al (2017) Meneco, a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks. PLOS Comput Biol 13:e1005276. doi: 10.1371/journal.pcbi.1005276

Querellou J, Cadoret J-P, Allen MJ, Collén J (2010) Marine Biotechnology. In: Cock J, Tessmar-Raible K, Boyen C, Viard F (eds) Introduction to Marine Genomics, vol 1. Springer, Netherlands, Dordrecht, pp 287–313

Rausch C, Hoof I, Weber T, et al (2007) Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. BMC Evol Biol 7:78. doi: 10.1186/1471-2148-7-78

Regueira TB, Kildegaard KR, Hansen BG, et al (2011) Molecular basis for mycophenolic acid biosynthesis in Penicillium brevicompactum. Appl Environ Microbiol 77:3035–3043. doi: 10.1128/AEM.03015-10

Rizzo JM, Buck MJ (2012) Key Principles and Clinical Applications of "Next-Generation" DNA Sequencing. Cancer Prev Res 5:887–900. doi: 10.1158/1940-6207.CAPR-11-0432

Robin J, Jakobsen M, Beyerm M, et al (2001) Physiological characterisation of Penicillium chrysogenum strains expressing the expandase gene from Streptomyces clavuligerus during batch cultivations. Growth and adipoyl-7-aminodeacetoxycephalosporanic acid production. Appl Microbiol Biotechnol 57:357–362. doi: 10.1007/s002530100787

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

Rojas-Aedo JF, Gil-Durán C, Del-Cid A, et al (2017) The Biosynthetic Gene Cluster for Andrastin A in Penicillium roqueforti. Front Microbiol 8:813. doi: 10.3389/fmicb.2017.00813

Romero D, Traxler MF, López D, Kolter R (2011) Antibiotics as Signal Molecules. Chem Rev 111:5492–5505. doi: 10.1021/cr2000509

Ropars J, Rodrıguez de la Vega RC, Lopez-Villavicencio M, et al (2015) Adaptive Horizontal Gene Transfers between Multiple Cheese-Associated Fungi. Curr Biol 25:2562–2569. doi: 10.1016/j.cub.2015.08.025

Roze L V., Chanda A, Linz JE (2011) Compartmentalization and molecular traffic in secondary metabolism: A new understanding of established cellular processes. Fungal Genet Biol 48:35–48. doi: 10.1016/j.fgb.2010.05.006

Roze L V, Chanda A, Laivenieks M, et al (2010) Volatile profiling reveals intracellular metabolic changes in Aspergillus parasiticus: veA regulates branched chain amino acid and ethanol metabolism. BMC Biochem 11:33. doi: 10.1186/1471-2091-11-33

Rugbjerg P, Naesby M, Mortensen UH, Frandsen RJN (2013) Reconstruction of the biosynthetic pathway for the core fungal polyketide scaffold rubrofusarin in Saccharomyces cerevisiae. Microb Cell Fact 12:31. doi: 10.1186/1475-2859-12-31

Sallam LA, El-Refai AM, Hamdi AH, et al (2005) Studies on the application of immobilization technique for the production of cyclosporin A by a local strain of Aspergillus terreus. J Gen Appl Microbiol 51:143–9.

Salo O V, Ries M, Medema MH, et al (2015) Genomic mutational analysis of the impact of the classical strain improvement program on β-lactam producing Penicillium chrysogenum. BMC Genomics 16:937. doi: 10.1186/s12864-015-2154-4

Sánchez BJ, Zhang C, Nilsson A, et al (2017) Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Mol Syst Biol 13:935.

Sardi M, Gasch AP (2017) Incorporating comparative genomics into the design–test–learn cycle of microbial strain engineering. FEMS Yeast Res. doi: 10.1093/femsyr/fox042

Schinko T, Berger H, Lee W, et al (2010) Transcriptome analysis of nitrate assimilation in Aspergillus nidulans reveals connections to nitric oxide metabolism. Mol Microbiol 78:720–738. doi: 10.1111/j.1365-2958.2010.07363.x

Shelest E (2008) Transcription factors in fungi. FEMS Microbiol Lett 286:145–151. doi: 10.1111/j.1574-6968.2008.01293.x

Sieber SA, Marahiel MA (2005) Molecular Mechanisms Underlying Nonribosomal Peptide Synthesis: Approaches to New Antibiotics. Chem Rev 105:715–738. doi: 10.1021/cr0301191

Siewers V, Chen X, Huang L, et al (2009) Heterologous production of non-ribosomal peptide LLD-ACV in Saccharomyces cerevisiae. Metab Eng 11:391–397. doi: 10.1016/j.ymben.2009.08.002

Smith DJ, Burnham MK, Bull JH, et al (1990) Beta-lactam antibiotic biosynthetic genes have been conserved in clusters in prokaryotes and eukaryotes. EMBO J 9:741–7.

Smith S, Tsai S-C (2007) The type I fatty acid and polyketide synthases: a tale of two megasynthases. Nat Prod Rep 24:1041–1072. doi: 10.1039/b603600g

Stassen PM, Kallenberg CGM, Stegeman CA (2007) Use of mycophenolic acid in non-transplant renal diseases. Nephrol Dial Transplant 22:1013–9. doi: 10.1093/ndt/gfl844

Takeda I, Myco U, Koike H, et al (2014) Motif-independent prediction of a secondary metabolism gene cluster using comparative genomics: Application to sequenced genomes of Aspergillus and ten other filamentous fungal species. DNA Res 21:447–457. doi: 10.1093/dnares/dsu010

Tang X, Li J, Millán-Aguiñaga N, et al (2015) Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-directed Genome Mining. ACS Chem Biol 10:2841–2849. doi: 10.1021/acschembio.5b00658

Tannous J, El Khoury R, Snini SP, et al (2014) Sequencing, physical organization and kinetic expression of the patulin biosynthetic gene cluster from Penicillium expansum. Int J Food Microbiol 189:51–60. doi: 10.1016/j.ijfoodmicro.2014.07.028

Tatusov RL, Fedorova ND, Jackson JD, et al (2003) The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:41. doi: 10.1186/1471-2105-4-41

Tatusov RL, Galperin MY, Natale DA, Koonin E V (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33–6.

Terabayashi Y, Shimizu M, Kitazume T, et al (2012) Conserved and specific responses to hypoxia in Aspergillus oryzae and Aspergillus nidulans determined by comparative transcriptomics. Appl Microbiol Biotechnol 93:305–317. doi: 10.1007/s00253-011-3767-4

Terrasan CRF, Temer B, Duarte MCT, Carmona EC (2010) Production of xylanolytic enzymes by Penicillium janczewskii. Bioresour Technol 101:4139–43. doi: 10.1016/j.biortech.2010.01.011

Thompson DA, Roy S, Chan M, et al (2013) Evolutionary principles of modular gene regulation in yeasts. Elife 2:1–37. doi: 10.7554/eLife.00603

Thykaer J, Nielsen J (2003) Metabolic engineering of beta-lactam production. Metab Eng 5:56–69. doi: 10.1016/S1096-7176(03)00003-X

Thykaer J, Rueksomtawin K, Noorman H, Nielsen J (2008) NADPH-dependent glutamate dehydrogenase in Penicillium chrysogenum is involved in regulation of b-lactam production. Microbiology 154:1242–1250. doi: 10.1099/mic.0.2007/010017-0

Umemura M, Koike H, Nagano N, et al (2013) MIDDAS-M: Motif-independent de novo detection of secondary metabolite gene clusters through the integration of genome sequencing and transcriptome data. PLoS One 8:e84028. doi: 10.1371/journal.pone.0084028

Valiante V, Mattern DJ, Schu A, et al (2017) Discovery of an Extended Austinoid Biosynthetic Pathway in Aspergillus calidoustus. ACS Chem Biol 12:1227–1234. doi: 10.1021/acschembio.7b00003

van den Berg MA, Albang R, Albermann K, et al (2008) Genome sequencing and analysis of the filamentous fungus Penicillium chrysogenum. Nat Biotechnol 26:1161–8. doi: 10.1038/nbt.1498

Väremo L, Nielsen J, Nookaew I (2013) Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. Nucleic Acids Res 41:4378–4391. doi: 10.1093/nar/gkt111

Vernikos G, Medini D, Riley DR, Tettelin H (2015) Ten years of pan-genome analyses. Curr Opin Microbiol 23:148–154. doi: 10.1016/j.mib.2014.11.016

Viggiano A, Salo O, Ali H, et al (2017) Elucidation of the biosynthetic pathway for the production of the pigment chrysogine by Penicillium chrysogenum. Appl Environ Microbiol AEM.02246-17. doi: 10.1128/AEM.02246-17

Visagie CM, Houbraken J, Frisvad JC, et al (2014) Identification and nomenclature of the genus Penicillium. Stud Mycol 78:343–71. doi: 10.1016/j.simyco.2014.09.001

Wang B, Yangyong L, Xuejie L, et al (2017) Profiling of secondary metabolite gene clusters regulated by LaeA in Aspergillus niger FGSC A1279 based on genome sequencing and transcriptome analysis. Res Microbiol. doi: 10.1016/J.RESMIC.2017.10.002

Weber T, Blin K, Duddela S, et al (2015) antiSMASH 3.0--a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 43:W237–W243. doi: 10.1093/nar/gkv437

Webster J, Weber R (2007) Introduction to fungi, 3rd editio. Cambridge University Press

WHO (2017) Antibiotic resistance. In: World Heal. Organ. http://www.who.int/mediacentre/factsheets/antibiotic-resistance/en/. Accessed 20 Dec 2017

Wiemann P, Guo C-J, Palmer JM, et al (2013) Prototype of an intertwined secondary-metabolite supercluster. Proc Natl Acad Sci U S A 110:17065–70. doi: 10.1073/pnas.1313258110

Wisecaver JH, Rokas A (2015) Fungal metabolic gene clusters-caravans traveling across genomes and environments. Front Microbiol 6:161. doi: 10.3389/fmicb.2015.00161

Wolf T, Shelest V, Nath N, Shelest E (2016) CASSIS and SMIPS: Promoter-based prediction of secondary metabolite gene clusters in eukaryotic genomes. Bioinformatics 32:1138–1143. doi: 10.1093/bioinformatics/btv713

Woo PCY, Lam C-W, Tam EWT, et al (2014) The biosynthetic pathway for a thousand-year-old natural food colorant and citrinin in Penicillium marneffei. Sci Rep 4:6728. doi: 10.1038/srep06728

Yadav G, Gokhale RS, Mohanty D (2009) Towards Prediction of Metabolic Products of Polyketide Synthases: An In Silico Analysis. PLoS Comput Biol 5:e1000351. doi: 10.1371/journal.pcbi.1000351

Yang Y, Zhao H, Barrero RA, et al (2014) Genome sequencing and analysis of the paclitaxel-producing endophytic fungus Penicillium aurantiogriseum NRRL 62431. BMC Genomics 15:69. doi: 10.1186/1471-2164-15-69

Yeh H-H, Ahuja M, Chiang Y-M, et al (2016) Resistance Gene-Guided Genome Mining: Serial Promoter Exchanges in *Aspergillus nidulans* Reveal the Biosynthetic Pathway for Fellutamide B, a Proteasome Inhibitor. ACS Chem Biol 11:2275–2284. doi: 10.1021/acschembio.6b00213

Yu D, Xu F, Zhang S, Zhan J (2017) Decoding and reprogramming fungal iterative nonribosomal peptide synthetases. Nat Commun 8:15349. doi: 10.1038/ncomms15349

Ziemert N, Alanjary M, Weber T (2016) The evolution of genome mining in microbes – a review. Nat Prod Rep 33:988–1005. doi: 10.1039/C6NP00025H

Ziemert N, Lechner A, Wietz M, et al (2014) Diversity and evolution of secondary metabolism in the marine actinomycete genus Salinispora. Proc Natl Acad Sci U S A 111:E1130-9. doi: 10.1073/pnas.1324161111