Nonribosomal peptides in fungal cell factories: from genome mining to optimized heterologous production

Antoine Vassaux, Loic Meunier, Micheline Vandenbol, Denis Baurain, Patrick P. Fickers, Philippe Jacques, Valérie Leclère

Nonribosomal peptides in fungal cell factories: from genome mining to optimized heterologous production

Antoine Vassaux[a,b], Loïc Meunier[a,c], Micheline Vandenbol[d], Denis Baurain[c], Patrick Fickers[a], Philippe Jacques[a], Valérie Leclère[b*]

[a]TERRA Teaching and Research Centre, Microbial Processes and Interactions, Gembloux Agro-Bio Tech, University of Liege, Avenue de la Faculté d'Agronomie, B5030 Gembloux, Belgium

[b]EA 7394-ICV-Institut Charles Viollette, Univ. Lille, INRA, ISA, Univ. Artois, Univ. Littoral Côte d'Opale, Avenue Paul Langevin, 59655 Villeneuve d'Ascq, France

[c]InBioS-PhytoSYSTEMS, Eukaryotic Phylogenomics, University of Liege, Boulevard du Rectorat 27, B-4000 Liège, Belgium

[d]TERRA Teaching and Research Centre, Microbiologie et Génomique, Gembloux Agro-Bio Tech, University of Liege, Avenue de la Faculté d'Agronomie, B5030 Gembloux, Belgium

*** Corresponding author:**
Pr. Valérie Leclère
e-mail: valerie.leclere@univ-lille.fr

Abstract

Fungi are notoriously prolific producers of secondary metabolites including nonribosomal peptides (NRPs). The structural complexity of NRPs grants them interesting activities such as antibiotic, anti-cancer, and anti-inflammatory properties. The discovery of these compounds with attractive activities can be achieved by using two approaches: either by screening samples originating from various environments for their biological activities, or by identifying the related clusters in genomic sequences thanks to bioinformatics tools. This genome mining approach has grown tremendously due to recent advances in genome sequencing, which have provided an incredible amount of genomic data from hundreds of microbial species. Regarding fungal organisms, the genomic data have revealed the presence of an unexpected number of putative NRP-related gene clusters. This highlights fungi as a goldmine for the discovery of putative novel bioactive compounds. Recent development of NRP dedicated bioinformatics tools have increased the capacity to identify these gene clusters and to deduce NRPs structures, speeding-up the screening process for novel metabolites discovery. Unfortunately, the newly identified compound is frequently not or poorly produced by native producers due to a lack of expression of the related genes cluster. A frequently employed strategy to increase production rates consists in transferring the related biosynthetic pathway in heterologous hosts. This review aims to provide a comprehensive overview about the topic of NRPs discovery, from gene cluster identification by genome mining to the heterologous production in fungal hosts. The main computational tools and methods for genome mining are herein presented with an emphasis on the particularities of the fungal systems. The different steps of the reconstitution of NRP biosynthetic pathway in heterologous fungal cell factories will be discussed, as well as the key factors to consider for maximizing productivity. Several examples will be developed to illustrate the potential of heterologous production to both discover uncharacterized novel compounds predicted *in silico* by genome mining, and to enhance the productivity of interesting bio-active natural products.

Keywords: Bioinformatics tools; Fungal cells; Genome mining; Heterologous production; Large gene cluster; Multi-modular enzyme; Nonribosomal peptide; Nonribosomal peptide synthetase; Secondary metabolite

Abbreviations

A-domain	Adenylation domain
ACV	d-($_{L}$-α-aminoadipyl)-$_{L}$-cysteinyl-$_{D}$-valine
BCGs	Biosynthetic gene clusters
C_{term}-domain	Terminal condensation domain
NRP	Nonribosomal peptide
NRPS	Nonribosomal peptides synthetase
ORF	Open reading frame
pHMMs	Profile hidden markov models
PKS	Polyketide synthase
PPTase	Phosphopantetheine transferase
SGS	Second generation sequencing
SMs	Secondary metabolites
TAR cloning	Transformation-associated recombination cloning
TGS	Third generation sequencing

1. Introduction

The era of nonribosomal peptides (NRPs) started with the brilliance of Fleming to analyze the unexpected ability of a filamentous *Penicillium* strain to impair bacterial growth. These observations led to the discovery of the penicillin antibiotic, a NRP derived compound which was not identified as such at that time (Fleming, 1929). Indeed, the biosynthesis mechanism of this type of peptides, named thiotemplate mechanism, was elucidated only 42 years later with the description of the Non Ribosomal Peptide Synthetase (NRPS) of the gramicidin antibiotic from *B. subtilis* (Lipmann et al., 1971). Since then, there has been growing interest in the study of this kind of compounds. As a result, NORINE, the first database entirely dedicated to NRPs was established in 2007, starting with 700 NRPs entries (Caboche et al., 2008). At present, over a thousand of NRPs have been characterized and classified (Pupin et al., 2016). These NRPs were mostly found in prokaryotes and fungal eukaryotes (Demain, 2014), and are notably not present in metazoa and plantae eukaryotic clades. Few exceptions were reported such as the production of astins in the plant *Aster tataricus*, latter refuted and attributed to a fungal endophyte *Cyanodermella asteris* (Schafhauser et al., 2019 - submitted). Likewise, polytheonamide initially isolated from the marine sponge *Theonella swinhoei* (Hamada et al., 2005), was finally shown to be produced by a symbiotic bacteria (Freeman et al., 2016; Wilson et al., 2014).

Nonribosomal peptides, as suggested by their name, are not assembled on the ribosome but by large multi-modular enzymatic complexes that are encoded by gene clusters that often span over tens of kilobases in the genome (**Fig. 1**). However, NRP synthetases are synthesized on the ribosome, before being activated by a specific phosphopantetheinyl transferase (PPTase). This PPTase is responsible for the conversion of the *apo*-NRPS into active *holo*-NRPS through the addition of a so-called phosphopantetheine arm onto thiolation domains (Beld et al., 2014). NRPSs can be divided into multiple modules, each of them responsible for the incorporation of one peptide monomer. These modules comprise multiple enzymatic domains exhibiting specific functions in the NRP assembly line (Caboche et al., 2008). The adenylation (A) domain is able to recognize and activate a specific monomer, which is then tethered onto a thiolation (T) / peptidyl carrier protein domain through the phosphopantetheine arm. The flexibility of the multi-enzymatic complex enables the presentation of monomers to the condensation (C) domain which catalyzes the formation of a peptide bond between monomers from two subsequent modules (Winn et al., 2016). In NRPSs involved in lipopeptide biosynthesis, a major NRPs class, a starter C domain priming the synthetase enables to acylate the first peptide monomer with the fatty acid part (Rausch et al., 2007). Some NRPSs also exhibit

additional domains responsible for monomer modifications, such as methylation, oxidation and epimerization.

NRPSs predominantly operate according to a linear synthesis, with an initiation module (A-T) able to recognize the first monomer, followed by as many modules (C-A-T) as monomers required to complete the peptide. Nonetheless, some NRPSs rather operate with an iterative or nonlinear organization, and then may not follow the colinearity rule between the number of modules and the size of synthesized peptides (Richter et al., 2014; Schwecke et al., 2006). The amonabactin synthetase found in *Aeromonas* sp. is a prime example of an atypical assembly line which displays an iterative, alternative and optional mode of synthesis. In this NRPS, the AmoG A-domain activates alternatively a Trp or a Phe, on which are bonded two dipeptides iteratively generated by the AmoE and AmoF domains, and optionally in two out of four cases, an AmoH domain incorporates a glycine between the 2,3-dihydroxy benzoic acid and the lysine of one of the dipeptides (Esmaeel et al., 2017). NRPS assembly lines usually end with a termination module comprising a thioesterase domain (TE) responsible for the release, and the cyclization, in most cases, of the final compound (Du and Lou, 2010; Kopp and Marahiel, 2007). In fungi, thioesterase domains can be replaced by condensation terminal domain (C_{term}) ending the NRPS and leading to cyclic NRPs (Gao et al., 2012). Alternatively, in NRPSs involved in peptaibols biosynthesis (i.e., NRPs harboring a C-terminal alcoholic group instead of carboxyl group), TE is substituted by a reductase domain releasing the peptide with a C-terminal alcohol (Manavalan et al., 2010).

Filamentous fungi are considered as prolific producers of NRPs. A large number of clusters involved in NRP synthesis has been detected for example in *Penicillium thymicola* (17 clusters), *Aspergillus fumigatus* (14 clusters), *Talaromyces islandicus*, *Beauveria bassiana* (13 clusters), and *Pestalotiopsis fici* (12 clusters) (Chiang et al., 2014; Gibson et al., 2014; Schafhauser et al., 2016; Tang et al., 2018; Wang et al., 2015). The high potential of fungal organisms for NRPs production can be exploited for the discovery of novel compounds of interest. The search for novel compounds in fungi is typically achieved following two different approaches: one starting from the biological activities and the other, more recent, from genome sequences. The principle of the former consists in screening samples originating from various environments for their biological activities. Once the targeted bioactivity is detected, the active natural product is purified and characterized (Kildgaard et al., 2017; Kunakom and Eustáquio, 2019). Usually bioactivities are first screened at small-scale and then upscaled to reach higher production rates. However, this strategy is very restrictive as only compounds displaying the screened activity and being produced in sufficient amount to induce an

observable activity, will be detected (Luo et al., 2014). In contrast, the second approach is based on the use of bioinformatics tools in order to identify the SM gene clusters of interest (Luo et al., 2014; Mushtaq et al., 2018). Frequently, the associated compound is not produced because the related cluster is cryptic and the cultivation conditions were probably not optimal to succeed in its awakening) (Anyaogu and Mortensen, 2015; Lazarus et al., 2014). Therefore, the second approach often requires genetic engineering either to activate the expression in the native host (homologous production) or to transfer the biosynthesis pathway in a heterologous host (heterologous production) (Luo et al., 2016).

In the past few years, microbial genome sequencing and annotation methods have made an incredible leap forward, providing the scientific community a wealth of publicly available fungal genomes (Grigoriev et al., 2014). This technological progress has opened new perspectives for the discovery of novel compounds through the bioinformatic approach. The development of novel bioinformatics tools enables, at present, the analysis of multi-modular complexes such as NRPSs, but also polyketide synthases (PKS) or hybrid complexes PKS-NRPS (Weber and Kim, 2016). The case of polyketides, assembled from acyl- and malonyl- coenzyme A by PKSs, is already well documented and is beyond the scope of this review (Chooi and Tang, 2012; Schümann and Hertweck, 2006; Zhang et al., 2011). Nonetheless, this study presents the multiple bioinformatics tools developed to predict the modular organization of NRPSs, and the substrate specificity of their adenylation domains, with an emphasis on the specificities of fungal systems. Bearing in mind that these NRPs identified by genome mining are frequently poorly synthesized under laboratory culture conditions, and taking into account for most of these fungi that adapted molecular tools are lacking, the heterologous expression of their biosynthetic pathways is often a necessary step (Nah et al., 2017; Sung et al., 2017; Tsunematsu et al., 2013a). The heterologous production of NRPs in bacterial organisms such as *Escherichia coli* and *Streptomyces* spp. have been widely reviewed (Beites and Mendes, 2015; Nah et al., 2017; Weber et al., 2015b; Zhang et al., 2011), but only few studies have investigated the potential of fungal organisms to produce such foreign compounds. Therefore, this review discusses, through multiple successful examples, the possibility to use fungal cell factories to heterologously express NRPS-encoding genes identified by genome mining. Finally, the multiple key factors to consider obtaining an efficient NRPs heterologous producer are described and then illustrated onto the example of the NRP-derivative penicillin, a multi-decades case of study.

2. Computational tools and methods for genome mining

Biosynthetic gene clusters (BGCs) involved in the synthesis of secondary metabolites (SMs) usually contain core synthase genes, regulatory and transport-related elements, and sometimes genes conferring a resistance to the related SM. Most of the developed bioinformatics tools can be applied to the analysis of BGCs related to multiple type of SMs (e.g. NRPS, PKS) coming out from either bacterial or fungal genomic data. Regarding to NRPS analysis, bioinformatics methods have been developed based on the high conservation of the NRPS thiotemplate mechanism and decades of experimental characterization (Weber, 2014; Ziemert et al., 2016). Applications of these algorithms can be used to identify NRPS cluster genes, their subsequent annotation, the prediction of the backbone of the putative NRP produced, and the search for related NRPs. The *in silico* annotation of BGCs relies on the algorithm efficiency, but also on the genome assembly quality, as the quality and the completeness of the annotation depend directly on the input sequences (Yandell and Ence, 2012). These two aspects will be further discussed in the following sections.

2.1 Recent advances in genome sequencing and annotation

From the mid-2000s, Second Generation Sequencing (SGS) technologies, also named Next Generation Sequencing (including Illumina and Ion Torrent technologies), revolutionized genomics and many other biological fields, by offering low-cost genome sequencing (Loman et al., 2012). These developments generated an exponential growth in the genome sequences available in public databases (more than 190,000 prokaryotic genome assemblies available on NCBI databases as of March 2019, https://www.ncbi.nlm.nih.gov/genome/browse/) (Tatusova et al., 2014). SGS technologies are based on the principle of a high-throughput method, with a short read length up to a few hundred bases (e.g., 250 bp for Illumina HiSeq), and usually a low error rate. The drawbacks of short read genome assemblies are related to low contiguity and poor gene completeness (Heather and Chain, 2016; Koren and Phillippy, 2015). As NRPS biosynthetic gene clusters can span several tens of thousands bases - even more when considering “accessory genes” -, BGCs full sequence might be spread into several contigs (Conway and Boddy, 2013; Vesth et al., 2016). This causes complex issues for reconstructing BGCs from draft genomes, as well as determining if several unknown gene clusters in an organism are related to only one or several NRPSs.

The recent emergence of Third Generation Sequencing (TGS), with PacBio and Oxford Nanopore technologies, brings new interesting features for BGC annotation. These two methods yield long read length (from a few tens of kb for PacBio, up to 1 Mb for Oxford Nanopore, depending on the quality of the DNA extraction (Jain et al., 2018)), hence generating high contiguity

assemblies (Lu et al., 2016). This long read length may present a real advantage in the case of BGC studies, allowing for obtaining complete sequences, and therefore the entire set of involved genes. Yet, in contrast to SGS, TGS displays a lower throughput and a higher error rate (Rhoads and Au, 2015). Several software packages, such as Flye (Kolmogorov et al., 2019) and Canu (Koren et al., 2017), provide efficient methods for TGS data assembly. When the read depth coverage is not sufficient, hybrid strategies taking advantage of the low error rate of SGS can be used for polishing TGS reads or contigs, allowing for obtaining a high-quality genome assembly (Bashir et al., 2012). For instance, SPAdes (Antipov et al., 2016) offers hybrid assemblies, while Pilon (Walker et al., 2014) allows the user to polish TGS assemblies.

Regarding the NRPS identification and annotation, these steps are driven by fast and reasonably accurate bioinformatics methods, which have now become one of the most predominant approaches to support the discovery of novel SMs. This emerging field is promising, as a result of constant effort in software development (Chavali and Rhee, 2018; Weber and Kim, 2016; Ziemert et al., 2016) and the availability of genomic data. Most methods for annotating NRPSs rely on similarity searches, which is a robust approach as these multi-enzymatic systems are composed of repeated, well conserved domains. The detection of core biosynthetic components is commonly conducted by protein signature searches (Chavali and Rhee, 2018). These signatures, also called pHMMs (profile Hidden Markov Models) are probabilistic models capturing the versatile information contained in a multiple sequence alignment (i.e., amino acid composition and insertions/deletions per site) to perform specific and sensible annotations (Eddy, 2011, 1998). Hence, from the prediction of synthetase genes, it is then possible to deduce the NRPS architecture (i.e., domain composition), and search for additional genes surrounding this core (e.g., regulation and transport related genes, tailoring enzymes).

2.2 Bioinformatics tools dedicated to NRPS study

Publicly available bioinformatics tools for annotating NRPS started to be devised a decade ago (Blin et al., 2017a), since then, a large number of software packages and methodologies for genome mining analyses has been released (**Fig. 2**). The Secondary Metabolite Bioinformatics Portal represents a very useful online resource (http://www.secondarymetabolites.org/) facilitating the exploration of these numerous software packages. It gathers information about the majority of genome mining and annotation tools, as well as databases applied to secondary metabolism investigation (Weber and Kim, 2016). The portion of these tools implying specific functionalities for

NRPS and closely related PKS is mostly constituted with generic algorithms applicable for both bacteria and fungi, even if the training of their algorithm was mainly based on bacterial data. For instance, NRPS annotation: NP.searcher (Li et al., 2009), ClustScan (Starcevic et al., 2008), and antiSMASH (Blin et al., 2017a) are generic tools for the overall annotation of NRPS. The latter is currently the most comprehensive and up-to-date pipeline for BGC analysis, providing a great deal of available functionalities for refining NRPS annotations (Blin et al., 2017a; Weber and Kim, 2016). Among the small portion of tools dedicated to Fungi, SMURF (Khaldi et al., 2010) is also a well-known software performing the basic identification of putative synthase and auxiliary enzyme genes. It is also noteworthy that antiSMASH possesses a version for fungal data (i.e., fungiSMASH - http://fungismash.secondarymetabolites.org) (Blin et al., 2017c) which adapts the internal gene predictor for an eukaryotic one (GlimmerHMM). In addition, fungiSMASH also offers to apply CASSIS (Wolf et al., 2016), an algorithm aiming the improvement of NRPS/PKS BGC boundaries delimitation by searching the presence of common regulation pattern in BGC promoters.

In addition to these general NRPS annotation tools, some are entirely devoted to the annotation of NRPS biosynthetic pathways, focusing on specific features of NRPSs and their products. Indeed, these tools aim to predict the peptidyl backbone of the putative NRP produced by analyzing two elements: substrate specificity of adenylation domains and condensation domain subtypes classes.

Substrate selection is based on the specificity of adenylation domains. Determining this specificity is a key element in the structural characterization of the putative synthesized compound. The specificity-conferring code, known as “Stachelhaus or NRPS code”, was the first prediction method based on the identity of ten amino acid residues located in the phenylalanine-binding pocket of PheA (adenylation domain of the first module of the gramicidin synthetase) (Stachelhaus et al., 1999). These critical residues were demonstrated to be involved in the substrate specificity as they define the binding-pocket recruiting specifically the various monomers. Therefore, with multiple alignments of the already known NRPS A-domains, it was possible to establish a code that correlate these residues to specifically recognized amino acid substrates (Challis et al., 2000). Therefrom, several substrate specificity prediction methods were developed such as the pHMM-based approach analyzing the specific protein signatures in the active site (Minowa et al., 2007). This method was integrated and further developed in NRPSsp (http://www.nrpssp.com/) and NRPS-PKS-substrate-predictor (http://www.cmbi.ru.nl/NRPS-PKS-substrate-predictor/) (Khayatt et al., 2013; Prieto et al., 2012). Another method, based on a Support Vector Machine approach, is exploited in NRPSpredictor2. This software package provides predictions according to the physicochemical

properties of the 34 residues in a radius of 8 angstrom around the PheA binding pocket (Rausch et al., 2005; Röttig et al., 2011). Substrate prediction obtained through the above-mentioned methods are usually accurate when dealing with bacterial data but are still not entirely reliable with fungal data. Three reasons can justify these limitations regarding substrate prediction : (1) low sequence conservation between A-domains (10-40% identity) (Agüero-Chapin et al., 2016), (2) the existence of different A-domain patterns for a same recognized amino acid (Khayatt et al., 2013), or on the contrary, (3) A-domains displaying a relaxed selection of more than one substrate (Belshaw et al., 1999; Esmaeel et al., 2016; Richter et al., 2014). Therefore, new strategies have been developed to address the still unresolved challenge of substrate specificity prediction. One recently developed approach, SEQL-NRPS (http://services.birc.au.dk/seql-nrps/), enables to predict substrate specificities based on an alignment-free method detecting occurrences of motifs in domains primary structure (Knudsen et al., 2016). Recently, prediCAT, a phylogenetic-based algorithm enabling to estimate the degree of predictability of A-domain, has been developed and integrated in SANDPUMA (Chevrette et al., 2017). SANDPUMA is a prediction framework combining three methods: (1) a retrained version of the Support Vector Machine models applied in NRPSpredictor2, (2) a search for exact matches from active site motifs as in the Stachelhaus et al. (1999) approach and (3) a pHMMs analysis like the Khayatt et al. (2013) approach. The last antiSMASH version (4.0) modified its previous substrate predictors (Stachelhaus code, Minowa's pHMMs and NRPSpredictor2) to implement SANDPUMA (Blin et al., 2017c).

Condensation domains also exhibit substrate specificity during the elongation reaction, as well as some possible additional activities such as cyclization or epimerization. These domains can be divided in different subtypes (**Table 1**). The identification of the C domain-subtype is a crucial element for inferring the modifications introduced into the amino acids constituting the final peptide (e.g., identifying $_{D}$-amino acids or peptide cyclization) (Caradec et al., 2014). Analysis of condensation domains can be achieved with the NaPDoS web application (http://napdos.ucsd.edu), which uses a phylogenetic approach to classify the subtypes of condensation domains and ketosynthase domains from PKS (Ziemert et al., 2012). Each subtype of condensation domain exhibits specific conserved motifs enabling to efficiently classify them in phylogenetic trees (Caradec et al., 2014). According to its position into the phylogenetic tree, it is possible by inference to determine the subtype of the analyzed C-domain.

When this additional information regarding core domains specificities has been collected, a quite precise determination of a putative NRP compounds architecture is possible by combining the

predicted substrates of adenylation domains, with the modifications related to the predicted condensation classes. Florine, a workflow dedicated to the annotation of NRP compounds, presents a detailed decision diagram to carry out this annotation process (Caradec et al., 2014). Nonetheless, the mechanism for substrate selection in NRPS is not fully elucidated and bioinformatics tools are being continuously improved as a result of new discoveries. For instance, the short amino acid regions connecting adjacent modules, called linkers, have been very recently highlighted to also play a role in the determination of substrate specificity. Subsequently to this discovery, a web application just has been released to enable the identification of these inter-modular NRPS linkers (https://nrps-linker.unc.edu) (Farag et al., 2019).

2.3 Linking NRPS to related clusters or known products

Once the NRPS architecture is elucidated thanks to the above described bioinformatics tools, it is important to determine if the compound is already described and associated to a characterized NRP (**Fig. 2**). A lot of useful tools are available to perform this dereplication steps, but only few of them will be described thereafter. The standard approach for analyzing BGC is to use the Basic Local Alignment Search Tool (BLAST) search program hosted on the NCBI website (http://www.ncbi.nlm.nih.gov/blast). This application enables to perform a sequence similarity search relative for instance to sequences listed in the "nucleotide" (BLASTn) or the "non-redundant protein sequences" (BLASTp) collections of the NCBI (McGinnis and Madden, 2004; NCBI Resource Coordinators, 2016). Alternatively, it is possible to query dedicated databases for BGCs, NRPSs and their products. Regarding BGC pathways, ClusterMine360 is a PKS/NRPS database mainly updated by crowdsourcing, which offers a two-way approach where nearly 1000 gene clusters can be browsed directly from a cluster list or following a search from more than 200 compound families (Conway and Boddy, 2013). The antiSMASH database is another relevant resource, encompassing currently a collection of antiSMASH results for almost 25,000 genomes - but mainly of bacterial origin - and offering the possibility to build precise queries (Blin et al., 2017b). MultiGeneBlast tool enables identifying homologs of multigene modules, such as NRPSs, on the basis of similarity and synteny criteria (Medema et al., 2013). MultiGeneBlast is also integrated in antiSMASH, where it suggests some potentially related clusters (or the identical cluster itself) taken from different BGC databases of varying levels of curation. For example, the "KnownClusterBlast" option refers to a curated database containing experimentally characterized clusters, while the "ClusterBlast" option uses an extended database composed mostly of only hypothetical clusters (Weber et al., 2015a).

The only database fully dedicated to NRP compounds, Norine, currently comprises 1190 peptides, including 346 NRPs of fungal origin distributed in 46 families. Norine offers different search modes: (1) an "annotation search" engine, which includes queries with all annotations (e.g., biological activity, molecular weight, monomer count, or producing organism), and (2) and a "structure mode" allowing searches based on monomer composition fingerprints, sequence similarity or whole structures. The analysis of predicted peptides permits searching for related analogous compounds already characterized, or to check if some amino acid patterns involved in some biological activities can be spotted (Pupin et al., 2016).. Only a few main databases were mentioned in this section, but a more exhaustive list is provided on the Secondary Metabolite Bioinformatics Portal (Weber and Kim, 2016).

2.4 Fungal NRPS analysis challenges

Most of the methods and bioinformatic tools available so far were designed both for the analysis of NRPS of fungal and bacterial origins. However, these organisms are taxonomically distant and some differences at the genomic level can be readily identified. Indeed, in comparison with prokaryotic genomes, fungal genomes are characterized by a larger size (i.e., several tens of millions bases), a higher number of chromosomes, a lower gene density, and the presence of introns (Elliott and Gregory, 2015; Setubal et al., 2018). BGC analysis pipelines, such as antiSMASH, offer an internal gene prediction before applying protein signature searches. Structural annotations (i.e., the prediction of whole gene structure) performed by automated internal gene predictors (e.g., Glimmer3 in antiSMASH (Medema et al., 2011)) are sufficiently accurate for bacteria, whereas annotation of genes from eukaryotic genomes is usually more complex. Dedicated gene prediction pipelines for eukaryotes often need some external data (gene expression and/or homologous protein data) to support *ab initio* predictions and provide higher quality annotation (Yandell and Ence, 2012). As it is essential to have coding sequences correctly annotated (i.e. with intron-exon and intergenic structures well resolved) to accurately predict NRPS gene clusters, fungal genomes are likely to benefit from a preliminary gene prediction, before being processed by BGC analysis tools. The available tools, dedicated to eukaryotic genome analysis, to perform this preliminary gene prediction are depicted in a relevant review (Yandell and Ence, 2012).

When pioneering predictive models for the substrate specificity of adenylation domains were developed a decade ago (Minowa et al., 2007; Rausch et al., 2005), very few fungal data was available. Therefore, these models provide predictions based mainly on monomers identified as a

substrate for the few well-characterized bacterial NRPS at that time. Consequently, these bioinformatics tools cannot predict with high accuracy the substrate specificity of most of the A-domains of eukaryotic NRPSs. Furthermore, based on the Norine database, it appears that NRPSs of fungal and bacterial origin display only a moderate range of overlapping substrate and incorporate a different set of monomers into the NRPs (**Fig. 3**). For example, Aib (2-aminoisobutyric acid) and amino alcohols such as Pheol (phenylalaninol) and Leuol (leucinol) seem to be specific to peptaibols produced mainly by fungi. Consequently, some fungal monomers are simply not recognized as no predictive models exist for these substrates. However, predictive models are only applicable if the monomer is directly selected by an adenylation domain and not resulting from the modification of a precursor amino acid. While Aib seems to be directly loaded (Chutrakul et al., 2008), amino alcohols in NRPs result from a modification of the amino precursors during synthesis of the peptide (Du and Lou, 2010; Manavalan et al., 2010). Determining if the monomer is loaded as such or modified during the synthesis is a complex task as no generalized rules can be applied. For example, D-amino acids are generally the result of an epimerization domain activity on L-amino acids, but, in some cases, D-Ala was reported to be directly loaded (McErlean et al., 2019).

Even if a few software packages, such as NRPSpredictor2 (Röttig et al., 2011), have previously developed models to predict the substrate specificity of adenylation domains, there is still an important need for devising tools enabling to predict more accurately which monomers are incorporated in fungal NRPs. Therefore, future challenges will be to go further in the elucidation of fungal NRPS mechanisms, and to establish novel tools allowing for increased the accuracy of substrate predictions in specific fungal systems, especially regarding to exotic- and D- amino acids.

3. Heterologous production of NRPS in fungal hosts

The number of available fungal genomic sequences in databases is continuously expanding thanks to individual or large-scale collaborative efforts, such as the “1000 fungal genomes project” (Araujo and Sampaio-Maia, 2018; Stajich, 2017). These available fungal genomes represent a gold mine for novel NRP compounds discovery. Unfortunately, in many cases, the newly sequenced fungi do not (or only poorly) produce these natural products under usual laboratory conditions in synthetic media, which are not always suitable to develop the filamentous fungi or to produce NRPs (Sung et al., 2017). One option is to activate or enhance the homologous production in the native host through promoter exchange for instance (Bode et al., 2015; Soukup et al., 2016). If this is not possible, due to a lack of molecular tools for example, the heterologous production is one of the best approaches to

reach a sufficient yield of production, and enable further characterization of the NRP (Alberti et al., 2017). In this part, the methodology commonly used for heterologous production of NRP is presented, taking into account different parameters such as the choice of the heterologous host, the genetic tools required for the transformation, the different expression systems used and the methods used for large scale heterologous production of NRP.

3.1 Heterologous host choice

The choice of the host is one of the most critical parameters to consider for heterologous gene expression and product synthesis. The selected strain has to present some desirable characteristics, such as accessible and efficient genetic manipulation tools, capabilities to grow at high cell density, ability to secrete the produced compounds, and above all, a codon compatibility with the foreign genes (Beites and Mendes, 2015). The well-known bacterial cellular platforms *Escherichia coli*, *Bacillus* sp., and *Streptomyces* sp., which display most of these characteristics, are commonly used for heterologous expression of NRPs. Nonetheless, bacteria exhibit a different codon usage from eukaryotic genes, which must be considered. One option for addressing this difficulty is to provide additional tRNA sequences for missing and rare codons, such as the AUA codon for Ile in *E. coli* (Skiba et al., 2018). Alternatively, as they are more closely related to filamentous fungi at a taxonomic level, yeasts constitute appropriate hosts to express foreign fungal NRPS genes. Different studies reporting NRP heterologous production in yeasts are presented in **Table 2**. Among yeasts, *Saccharomyces cerevisiae* displays both a high codon usage compatibility towards foreign fungal genes (Tsunematsu et al., 2013b), and a great potential for the secretion of the produced compounds (Mattanovich et al., 2012). This host also displays a certain tolerance towards foreign bacterial sequences, as was recently demonstrated with the heterologous production of a NRP initially produced by S*treptomyces lavendulae* (Wehrs et al., 2018). As it is possible in this yeast to express a NRPS-encoding gene coming from bacteria, which is taxonomically very distant to the host, its potential is even more promising for the heterologous production of fungal NRP. Moreover, the poorly developed secondary metabolism of *S. cerevisiae* is an asset, as it minimizes the competition with homologous SM pathways, and promotes the detection/purification of the targeted compound. Genetic manipulations are also made straightforward by a plethora of molecular tools, and the aptitude of the yeast to efficiently perform transformation-associated recombination (TAR) enabling the cloning of large DNA fragments (Anyaogu and Mortensen, 2015). However, *S. cerevisiae* lacks a functional PPTase encoding gene, which thus must be introduced into the host (Barajas et al., 2017).

In accordance with its advantages, and despite the risk of obtaining a low rate of production, *S. cerevisiae* is by far the most widespread yeast host for NRP heterologous production. Nevertheless, the non-conventional yeast *Hansenula polymorpha* (also belonging to the Saccharomycetales) can be an alternative choice. This methylotrophic yeast is also easily cultivable, even at large-scale, and strongly regulated promoters are available (Saraya et al., 2014). Moreover, *H. polymorpha* was successfully used to produce the NRP d-($_{L}$-α-aminoadipyl)-$_{L}$-cysteinyl-$_{D}$-valine (ACV), a precursor of penicillin and cephalosporin (Gidijala et al., 2009). Another Saccharomycetales, *Pichia pastoris* (now *Komagatella pastoris*) is also an appropriate host to heterologously express large SM clusters. In this host, multiple examples of large multi-modular PKS expression have been reported (Kasahara et al., 2010; Xue et al., 2017). Nonetheless, to date, no NRP heterologous production has been reported, despite the obvious potential of *P. pastoris* for SM production. Finally, *Schizosaccharomyces pombe* is the only yeast reported to produce naturally a NRP (ferrichrome), which indicates that it expresses a functional PPTase (Bushley et al., 2008; Schwecke et al., 2006). This characteristic would establish this yeast as a suitable host, but the limited genetic toolbox for Schizosaccharomycetales (only distantly related to Saccharomycetales such as *S. cerevisiae*) currently restricts its scope of usage (Erler et al., 2006).

Despite yeasts different filamentous fungi, taxonomically close to the original host, can be used for the heterologous expression of foreign NRPS genes. Numerous filamentous fungi display interesting intrinsic properties that might make them potentially convenient hosts for heterologous production of NRPs. First of all, they display a highly developed secondary metabolism, with the expression of a large number of NRPs/PKs, and therefore contain a functional PPTase. In addition, filamentous fungi were numerous times demonstrated to be able to recognize and correctly splice introns from foreign fungal mRNA sequences, bypassing the necessity to purify and clone large introns-free cDNAs from NRP gene cluster (Lazarus et al., 2014; Unkles et al., 2014). Usually, the choice of a taxonomically closely-related host is preferable, since it generally exhibits a good compatibility regarding codon usage, thus facilitating translation (Zhang and Liu, 2016). Indeed, in native hosts, mRNA sequences harbour mainly preferred codons, interspersed with scarcer codons in specific zones, so as to slow down the ribosome and enable the pre-folding of the translated proteins (Ongley et al., 2013). Among the potential fungal hosts, *Aspergillus* spp., such as *A. nidulans*, *A. niger*, and *A. oryzae*, are the more widely used for NRP heterologous production (**Table 3**). These species are relatively easy to grow on cheap carbon sources in submerged cultures. Moreover, they have been already used for decades at industrial-scale for food and pharmaceutical applications

(Mattern et al., 2015; Pickens et al., 2014). Conversely, *Aspergillus* spp. exhibit an extensively developed secondary metabolism, which may constitute a drawback. Indeed, the subsequent detection and purification of heterologous compounds is complicated due to the presence of a multitude of natural products. Among those species, *A. oryzae* displays the most limited endogenous secondary metabolism, providing a trade-off to the drawback just described (He et al., 2018). Moreover, this species is GRAS (generally recognized as safe), especially as concerns the absence of production of aflatoxin, a common mycotoxin found in most of *Aspergillus* species (Sakai et al., 2008). Lastly, numerous genetic tools, such as selectable markers and various strong promoters, are available to easily manipulate *A. oryzae* (Pahirulzaman et al., 2012). This is partly because of its close relatedness with the genetic model *A. nidulans*, thus resulting in the compatibility of genetic tools (Alberti et al., 2017).

3.2 Genetic toolbox for the cloning of genes involved in NRP biosynthesis

Once an appropriate host has been selected, the genes involved in the biosynthesis of the NRP have to be transferred from the natural producer to the heterologous host. In the case of NRPs heterologous production, this step can be particularly challenging for two reasons: (1) NRPS-encoding genes can encompass tens of kilobases which are not easily amplifiable by classical PCR operation, (2) NRPs synthesis often relies not only on the expression of the NRPS encoding-gene but required to transfer a set of genes which are generally grouped within the same cluster in the genome of the natural producer. This section will depict some of the solutions to overcome these issues.

3.2.1 Cloning of NRPS long sequences

As cloning NRPS genes encompassing up to tens of kilobases can be challenging, several methods have been developed to enable the reconstitution of a long fragment from multiple amplified PCR fragments (e.g., overlap extension PCR, Gibson assembly, USER cloning,) (**Fig. 4**). It is advisable to first perform a reverse transcription to generate a single-strand NRPS cDNA, which comprise the intron-free open reading frame, from a pool of total mRNA. The overlap extension PCR approach (**Fig. 4 -A**) permits then the assembly of multiple overlapping fragments, obtained via standard PCR, into a full length double-strand NRPS DNA (Bryksin and Matsumura, 2010; Horton et al., 1989). The Gibson assembly method (**Fig. 4 -B**) is another approach enabling the fusion of multiple DNA fragments. To set-up this method, the multiple PCR fragment are first amplified by standard PCR with primers displaying homologous sequences (approximately 40 bp). The isothermal one-step assembly of the fragments exhibiting terminal overlapping sequences is possible through the

joint action of a proofreading DNA polymerase, a Taq DNA ligase and a 5' T5 exonuclease. First, the 5' T5 exonuclease generates single-stranded 3' DNA overhangs by withdrawing nucleotides from the 5' ends. The generated homologous overhangs can then anneal, and the full-length double-stranded DNA be synthesized through the action of the DNA polymerase and the Taq DNA ligase. This system can generate very large DNA fragments, up to several hundreds of kilobases, with high efficiency (Gibson et al., 2009). Likewise, the Uracil-Specific Excision Reagent (USER) cloning (**Fig. 4 -C**) is also an efficient approach to assemble multiple DNA fragments. To implement USER, a deoxyuridine residue (U) is first included through the designed primers in each PCR fragment. Subsequently, the U residue is cleaved by a DNA glycosylase-lyase Endo VIII and a uracil DNA glycosylase, thus generating a 3' overhangs, enabling a directional assembly of PCR fragments in a linearized target plasmid (Cavaleiro et al., 2015; Smith et al., 1993).

Once the long fragment is isolated (e.g. gDNA, cDNA) or reconstituted through one of the above-mentioned approaches, TAR cloning is the most commonly used method for its transfer into cosmids or plasmids. Yeasts, especially *S. cerevisiae*, have the ability to perform homologous recombination resulting in the incorporation of the long fragment into the vector. The TAR cloning approach (**Fig. 4 -D**) implies first the cloning of two fragments homologous to the 5' and 3' parts of the long-length sequence, separated by a restriction site, into the vector. The vector comprising 5'-3' parts is then linearized on the above-mentioned restriction site and co-transformed in yeast with the full-length fragment (Kouprina and Larionov, 2016). The denomination "direct TAR Cloning" is used when gDNA or cDNA, eventually prior enzymatically digested, are directly used as a full-length fragment in the co-transformation (Zhang et al., 2019). On the contrary, if the full-length fragment was assembled by overlap extension PCR before to be inserted in the vector through homologous recombination, the method is named ExRec (overlap extension PCR – yeast homologous recombination) (Schimming et al., 2014; Tsunematsu et al., 2013b).

However, the insertion of long NRPS sequences into plasmids lead to a raise in the overall plasmid size, which is dramatically affecting the subsequent cloning efficiency. The maximum insert size can vary from one plasmid to another, but it is generally accepted that the insert size can reach up to 25 kb. Beyond this size, the use of cosmids, which are plasmids containing a phage sequence, enables to clone larger sequences (30-50 kb) in a more stable way (Carter and Shieh, 2015; Collins and Brüning, 1978). In filamentous fungi, an alternative method for cloning large SM clusters via Fungal Artificial Chromosomes (FACs) has been developed (Bok et al., 2015; Clevenger et al., 2017). An equivalent artificial chromosome system is also available in yeasts (YACs) (Hughes et al.,

2015). The use of artificial chromosomes has enabled the direct cloning of DNA fragments up to 300 kb (Zhang et al., 2019).

Alternatively, a very efficient way to isolate a pre-identified NRPS gene is to take advantage of *de novo* synthesis technology to re-assemble the constituting oligos. This method was already applied to synthesize the bottromycin NRPS gene with the purpose of expressing it in *Streptomyces* species (Huo et al., 2012), and could be similarly implemented in a fungal host. Nonetheless, even if the cost of this technology is continuously decreasing, it remains the most expensive way to obtain a long DNA sequence (Kosuri and Church, 2014).

3.2.2 Cloning of a large number of genes

Beside NRPS genes, numerous genes such as those encoding a PPTase, tailoring enzymes, or transporter-associated genes, are required to successfully produce a bioactive NRP or NRP derivative. The cloning of many genes can imply the synthesis of several plasmids which is laborious, costly, and problematic in case only few selection markers are available for the chosen host (Kumakura et al., 2019; Li et al., 2017). Several solutions particularly well-suited to NRP heterologous production have been proposed to alleviate this issue. One possible strategy consists of assembling the genes in the same plasmid to create a multiple-gene expression plasmid (Tsunematsu et al., 2013b). This approach was successively applied to heterologously express the aspyridone PKS-NRPS gene and its tailoring enzymes in *A. oryzae*. The aspyridone cluster identified in *A. nidulans* comprises the *adpA* PKS-NRPS gene and six genes encoding tailoring enzymes (*adpB*, *adpC*, *adpD*, *adpE*, *adpF*, and *adpG*). A plasmid was synthesized to enable the simultaneous expression of combinations of four genes under the control of four promoters (P_{amyB}, P_{adh}, P_{gpdA} and P_{eno}) in a single plasmid (Wasil et al., 2013). The golden gate cloning method can facilitate to a considerable extent the process of cloning multiple transcription units (promoter – gene – terminator) in a single expression vector. This strategy relies on the use of type IIs restriction enzymes that can generate a specific 5′ or 3′ DNA overhangs of 4 nucleotides depending of the enzyme considered. Subsequently to the enzymatic digestion, two fragments with appropriate cleavage site can be ligated in a one-step restriction-ligation (Engler et al., 2008). Based on the golden gate method, easily tunable vectors have been designed with proper cleavage sites enabling to introduce and interchange DNA elements (e.g. promoters, coding sequences, terminators, protein tag sequences, resistance genes) as mere building blocks. These kinds of DNA assembling platforms have already been designed for several potential host such as the filamentous fungus *Ustilago maydis* (Terfrüchte et al., 2014), and the yeasts

S. cerevisiae (Agmon et al., 2015; Mitchell et al., 2015), *S. pombe* (Kakui et al., 2015), and *Y. lipolytica* (Celińska et al., 2017). Alternatively, viral 2A peptide sequences can also be used to express several genes under the control of the same promoter in a single plasmid. These sequences, interspaced between each ORF, are cleaved during protein translation, hence releasing the proteins from each other (Kim et al., 2011; Unkles et al., 2014).

However, if the use of multiple plasmids and genomic integrations is required, methodologies allowing for recycling of selection markers have been developed in some common fungal hosts. The loss of *URA3* selection gene in *S. cerevisiae,* can be achieved by cultivation on 5-fluoroorotic acid plates (Akada et al., 2006; Boeke et al., 1987). The cre-lox system also enables selection marker recycling in yeast such as *S. cerevisiae* and *P. pastoris* (Pan et al., 2011; Sauer, 1987). A recycling system is also operational in *A. nidulans*, where it is possible to replace the marker of a previous transformation by cloning into the first selection gene, a second construction with a different marker (van Dijk and Wang, 2016; Yaegashi et al., 2014).

3.3 Fungal expression systems

3.3.1 Yeast heterologous expression systems

The first study reporting a successful heterologous expression of a fungal NRPS in yeast was related to the production of ACV in *H. polymorpha* (Gidijala et al., 2008). This example will be discussed in detail in section 4. In contrast, for all subsequently reported heterologous expressions of NRPS in yeast, *S. cerevisiae* was selected as a host (**Table 2**), for example with the fumitremorgin biosynthetic pathway from *A. fumigatus* (Tsunematsu et al., 2013a). The precursor of fumitremorgin is the NRP brevianamide F, resulting from a condensation of L-tryptophan and L-proline by FtmA NRPS. This compound is then prenylated by FtmB to form tryprostatin B, before being further modified by FtmC, FtmD, FtmE, and FtmF leading to spirotryprostatin A or B. The adopted approach for the cloning of the 8-kb *ftmA* gene was to take advantage of the capability of *S. cerevisiae* to perform TAR cloning (Kouprina and Larionov, 2016). With the purpose of removing introns, sometimes not or mis-spliced by heterologous yeasts, the total mRNA from the native producer was extracted and converted into cDNAs by a reverse transcriptase. After homologous recombination, the intron-free *ftmA* was inserted into a plasmid designed for expression in *S. cerevisiae* under the control of the strong galactose-inducible GAL1 promoter. After the cloning of *ftmB*, the NRP derivative tryprostatin B was detected at 100-fold higher concentration than in the native producer *A. fumigatus* (35.6 and 0.383 $mg.l^{-1}$, respectively) (Tsunematsu et al., 2013a). This

559 significant increase in tryprostatin B titers is a prime example for illustrating the potential of the
560 heterologous expression approach to improve NRP productivity.

561 A heterologous expression approach has also been used to experimentally assign an NRPS
562 cluster identified on the basis of genome mining to the synthesis of a given NRP. In *A. fumigatus*, 14
563 gene clusters were identified to be involved in the biosynthesis of NRP compounds (Frisvad et al.,
564 2009). Among them, Afu6g12080 was identified and predicted to be involved in fumiquinazoline
565 biosynthesis (Ames et al., 2010). Using a similar approach to that taken with *ftmA*, total mRNA was
566 extracted from *A. fumigatus* and converted into cDNAs by a reverse transcriptase. From this cDNA,
567 multiple overlapping PCR fragments with homologous ends (<3 kb) of the introns-free Afu6g12080
568 NRPS gene were amplified and assembled together using overlap extension PCR methodology. The
569 resulting 11.8-kb fragment was cloned under the control of the strong constitutive P_{tef1} promoter in
570 pKW1810, a *S. cerevisiae* expression plasmid, by TAR cloning. The coding sequences for Flag and
571 6xHis were added, respectively, to the 5' and 3' ends of the NRPS gene. The presence of tags attached
572 to protein ends facilitates the further detection and purification of the produced NRPS. In this study,
573 the authors were able to confirm the expression of Afu6g12080 by Western blot using antibodies
574 directed against 6xHis and Flag tags. addition of the gene *npgA* encoding a PPTase, fumiquinazoline
575 F (Anthranilic acid [Ant] - $_L$Trp - $_L$Ala), was purified from the culture broth of the constructed strain
576 (0.4 $mg.l^{-1}$). Moreover, an analog compound containing a hydroxylate-Ant was also identified,
577 highlighting the plasticity of the first A-domain of the NRPS encoded by Afu6g12080 (Ishiuchi et al.,
578 2012).

579 The Afu6g12080 NRPS from *A. fumigatus* is very similar to TqaA, another trimodular NRPS
580 identified in *Penicillium aethiopicum*. With the purpose of devising a ready-to-use host for
581 NRPS/PKS heterologous production, Gao and coworkers have introduced the *npgA* PPTase encoding
582 gene into the genome of *S. cerevisiae* (Strain BJ5464-NpgA). Meanwhile the *tqaA* cDNA, devoid of
583 introns, was obtained by reverse transcription on mRNA extracted from *P. aethiopicum*.
584 Subsequently, TAR cloning enabled the integration of the 5'-Flag modified NRPS gene in the yeast
585 expression plasmid pXW55 under the control of the strong constitutive ADH2 promoter. The
586 heterologous expression of *tqaA* was demonstrated through successful purification of the 450-kDa
587 synthetase with anti-Flag antibody affinity chromatography followed by gel filtration. Afterward, the
588 activity of the synthetase was confirmed with the detection of fumiquinazoline (Gao et al., 2012).

589 The strategy developed to express *tqaA* (Gao et al., 2012) was repeated several times to
590 heterologously express NRPS genes in the BJ5464-NpgA strain with the same plasmid. The *ardA*
591 NRPS gene from *Neosartorya fischeri*, was in such manner expressed in this strain, leading to
592 tricyclic ardeemin FQ production (Haynes et al., 2013). From the same filamentous fungi, *anaPS*,
593 involved in the biosynthesis of a precursor of acetylaszonalenin, was heterologously expressed to
594 investigate the role of the terminal condensation domain in NRP cyclization (Gao et al., 2012). The
595 expression of the 276-kDa bimodular synthetase AspA from *Aspergillus alliaceus* in *S. cerevisiae*
596 revealed the iterative mode of synthesis of the first module as it led to the synthesis of a tri-peptide
597 Ant-Ant-Trp (Gao et al., 2013). Likewise, the production of fungal anticancer cyclooligomer
598 depsipeptides, initially produced by *B. bassiana*, was reconstituted in *S. cerevisiae* through cloning of
599 the gene encoding the bassianolide synthetase (348 kDa) and the beauvericin synthetase (352 kDa)
600 (Yu et al., 2017, 2013).

601 Despite the opportunities offered by TAR cloning methodology, the insertion and correct
602 expression of large sequences, such as NRPS genes, remain difficult in numerous cases. One
603 potential approach to address this challenge is to dissect and clone the synthetase gene into separate
604 modules. The feasibility of module dissection was first been highlighted with the aspyridone
605 synthetase AdpA, encompassing a single NRPS module fused to the C-terminal part of a PKS. The
606 sequence encoding this module was isolated from *A. nidulans* and expressed in *S. cerevisiae*.
607 Subsequently the heterologous AdpA module was purified, prior to confirming the correct enzymatic
608 activity by *in vitro* production of acyltetramic acid in co-incubation with the PKS part (Xu et al.,
609 2010). Conversely, several mono-module NRPS-encoding genes from *Aspergillus terreus* were
610 expressed in the multi-copy pESC-URA plasmid: *apvA-melA*, *pgnA*, and ATEG_03090, respectively
611 involved in the biosynthesis of aspulyinone E, phenguignardic acid, and atromentin (Hühner et al.,
612 2018). The cryptic mono-module NRPS encoding gene ATEG_03630 from the same filamentous
613 fungi was similarly awakened by heterologous production, leading to the production of 2,4-
614 dihydroxy-5,6-dimethyl benzaldehyde (Wang and Zhao, 2014).

615 NRPSs encompassing more than one module can be dissected and each module cloned
616 separately. Nonetheless, some key factors must be considered. Primarily, the position where the
617 NRPS can be dissected is crucial. In the case of the bimodular AspA, the A_1-T_1-C_2 tridomain can be
618 solubly expressed unlike the bidomain A_1-T_1. Similarly, the second module was soluble in the four
619 domains C_2-A_2-T_2-Ct form and not in the A_2-T_2-Ct form (Gao et al., 2013). Recently, Bozhüyük et al.
620 (2019) have suggested an alternative fusion site area between the two sub-domains of the C-domain,

where NRPSs can be dissected. Authors postulated that the self-contained catalytically active unit can be redefined as C_{Asub}-A-T-C_{Dsub}, where C_{Asub} is the sub-domain accepting the growing peptide chain in the N-terminal C-domain and C_{Dsub} is the sub-domain providing the peptide chain in the C-terminal C-domain. Genetic engineering enables, by exchanging these functional units between different NRPSs with similar domain composition, to generate chimeric NRPSs assembling novel NRPs. Up to now, these units have been employed only to reconstitute entire chimeric NRPSs, and there is no evidence that they can be expressed separately in a soluble form (Bozhüyük et al., 2019). A second point to consider for the dissection of NRPSs is related to the presence of intermodular linkers, which are short amino acid regions connecting covalently subsequent modules. The importance of these short sequences has recently been highlighted between modules one and two of the beauvericin and bassianolide synthetases (Yu et al., 2013). The dissection of these synthetases by cloning module 1 and module 2+3 into two distinct plasmids without disrupting the pre-identified linker sequence led to an effective production of beauvericin and bassianolide. In contrast, the same experiment without the linker did not lead to any production of cyclooligomer depsipeptides. In addition to linkers, there are also short amino acid sequences, called COM domains, enabling noncovalent association of the different proteins when the NRPS is composed of several proteins such as in the tyrocidine synthetase or the surfactin synthase. These COM domains are essential when different fractions of the NRPS are expressed separately but can also be exploited to generate chimeric NRPS systems. This was demonstrated with the co-expression of the first module of the tyrocidine NRPS (TycA) from *Brevibacillus parabrevis* with the third module of the surfactin synthase (SrfAC) from *Bacillus subtilis* (Siewers et al., 2010). TycA was expressed with its native COM^D domain, whereas SrfAC was fused with the COM^A domain, compatible with COM^D, isolated from TycB module. Co-expression in *S. cerevisiae* of these two complexes into distinct pESC plasmids lead to the expected production of $_D$Phe-$_L$Leu.

3.3.2 Filamentous fungi heterologous expression systems

With the exception of the *pcbAB* NRPS gene (involved in ACV synthesis), expressed in *Neurospora crassa* (Smith et al., 1990), *Aspergillus* spp. were predominantly used to heterologously express NRP compounds (**Table 3**). One of the earliest reported NRP heterologous expressions in filamentous fungi was related to brevianamide F from *A. fumigatus*. The brevianamide NRPS gene, also later expressed in *S. cerevisiae* as described above, had previously been identified through heterologous expression in *A. nidulans* (Maiya et al., 2006). In order to prevent mis-amplification,

652 the 8.7-kb *ftmA* gene encoding the bimodular NRPS was amplified from *A. fumigatus* as two distinct
653 fragments (5' and 3'). The fragments were assembled together, cloned under the control of the alcohol
654 dehydrogenase promoter (P_{alcA}), and integrated in the genome of the host strain using *pyrG* as a
655 selection marker. In *A. nidulans,* P_{alcA} is a very strong promoter induced by ethanol or L-threonine,
656 and repressed in the presence of glucose (Felenbok et al., 2004). Cultivation under induction
657 conditions led to brevianamide F synthesis with a titer of 45 $mg.l^{-1}$, which represents a 3,500 fold
658 increase as compared to native strain *A. fumigatus* Af2936. Later, *ftmA* gene, assembled through
659 ExRec overlap-extension PCR-yeast homologous recombination (**Fig. 4**), was also expressed in *A.*
660 *niger* (Tsunematsu et al., 2013a). The FtmA synthetase from *N. fischeri*, displaying 88% identity at
661 the amino acids level with that from *A. fumigatus*, was also successfully expressed in *A. nidulans*
662 (Wunsch et al., 2015). *FtmA* gene under the control of the constitutive glyceraldehyde-3-phosphate
663 dehydrogenase promoter (P_{gpdA}) was inserted into the genome of *A. nidulans*, resulting in a
664 production of brevianamide F ranging between 9.7 and 36.9 $mg.l^{-1}$, depending on the mutant
665 considered. Interestingly, the productivity from one positive mutant to another displayed large
666 variations (from 1 to 4 magnitudes) (Maiya et al., 2006; Wunsch et al., 2015). This can be due to the
667 ectopic integration of the gene in variable regions of the genome.

668 Numerous fungal heterologous expression studies were conducted to decipher the mode of
669 synthesis of NRPSs. For example, the typical mode-of-operation of two siderophore synthetases has
670 been highlighted by heterologous expression in filamentous fungi. For this purpose, *FSN1* from
671 *Fusarium sacchari* and *CsNPS2* from *Ceriporiopsis subvermispora* were respectively expressed in *A.*
672 *oryzae* and *A. niger* (Brandenburger et al., 2017; Munawar et al., 2013). The *FSN1* gene is encoding
673 a NRPS involved in the synthesis of ferrirhodin, whereas *CsNPS2* is related to the biosynthesis of a
674 type VI siderophore. In both cases, the NRPSs were operating through an iterative mechanism with
675 three additional thiolation-condensation domains receiving the same monomer (N_5-acetyl-N_5-
676 hydroxy-L-ornithine) from a single upstream located adenylation domain. Likewise, through
677 heterologous expression in *A. nidulans*, the synthesis of asperphenamate has been demonstrated to be
678 dependent on the joint action of two distinct NRPSs: ApmA and ApmB. Indeed, the NRP produced
679 by ApmA, N-benzoylphenylalaninol, binds onto the monomer attached to ApmB to generate
680 subsequently the asperphenamate compound (Li et al., 2018). On the other hand, there are some
681 studies reporting identification and confirmation of NRP biosynthetic pathways through heterologous
682 expression. In the case of gliotoxin, a toxin synthesized by *A. fumigatus*, the related gene cluster was
683 known but the role of each gene had not been elucidated (Chang et al., 2013). NRPS gene *gliP*

684 expressed under the control of P_{alcA} in *A. nidulans* was identified to be involved in the first step of
685 gliotoxin biosynthesis with the assembly of NRP-type precursors: cyclo-($_{L}$-phenylalanyl-$_{L}$-seryl) and
686 cyclo-($_{L}$-tryptophanyl-$_{L}$-seryl). Likewise, by heterologous expression in *A. oryzae*, *pscyA* and *pscyB*
687 genes from *Penicillium sopii* have been demonstrated to be the only two genes required for the
688 biosynthesis of cycloaspeptide. Methylated amino acids constituting this NRP are provided through
689 the activity of an N-methyltransferase encoded by *pscyA*, while the peptide assembly is performed by
690 a 5-modules NRPS encoded by *pscyB* (De Mattos-Shipley et al., 2018). Another reported example
691 was the case of KK1, a NRP natively produced by *Curvularia clavata*, displaying anti-fungal activity
692 against the plant pathogen *Botrytis cinerea*. The 40-kb gene cluster potentially involved in KK1
693 synthesis was split into two fragments and cloned into *A. oryzae*. The transformed strain was able to
694 produce KK1, demonstrating the role of this gene cluster in the NRP synthesis (Yoshimi et al., 2018).

695 Some mono-modular NRPSs have also been heterologously expressed in filamentous fungi.
696 The *melA* gene from *A. terreus* was cloned into *A. nidulans* (Guo et al., 2015) and *A. niger* (Geib et
697 al., 2016; Geib and Brock, 2017) resulting in the production of aspulvinone E. Similarly, following
698 the cloning of *micA* obtained from *A. nidulans*, the transformed *A. niger* strain was able to produce
699 microperfuranone (Yeh et al., 2012).

700 *A. niger* was also used to heterologously express several cyclodepsipeptides, high value NRP
701 compounds, with anticancer, antimicrobial and anthelmintic properties (Sivanathan and
702 Scherkenbeck, 2014). Enniatin, beauvericin, and bassianolide belonging to this group of compounds
703 are assembled by iterative bimodular synthetases able to link $_{D}$-α-hydroxy acid with $_{L}$- amino acids.
704 The *esyn1* NRPS gene from *Fusarium oxysporum* was expressed in *A. niger,* leading to the
705 production of enniatin at initial un-optimized titer of 1 mg.l^{-1} (Richter et al., 2014). Beauvericin and
706 bassianolide were also heterologously produced after the cloning of the corresponding *bbBeas* and
707 *bbBsls* synthetase genes from *B. bassiana* (Boecker et al., 2018). In these two studies, a promising *A.*
708 *niger* heterologous expression host was developed based on the Tet-on bacterial-fungal hybrid
709 expression system. Tet-on system encompassed an rtTA2S-M2-dependent promoter controlling the
710 expression of the NRPS gene and a tetracycline-dependent transactivator rtTA2S-M2. Addition of a
711 tetracycline derivative, doxycycline (Dox), in the medium activated the system and induced the
712 expression of the gene. The expression strength is directly correlated to the Dox concentration added,
713 thus making Tet-on tuneable and very efficient, with expression levels equivalent to those obtained
714 with the widely used strong constitutive promoter P_{gpdA} (Meyer et al., 2011). In addition, the impact
715 of the number of *bbBeas* gene copies integrated in *A. niger* genome on the beauvericin productivity

716 was also investigated (Boecker et al., 2018). The strain harbouring a tandem gene copy produced
717 approximately two-fold higher titers of beauvericin than the strain harbouring a single copy, under
718 similar culture conditions (628 versus 294 $mg.l^{-1}$).

719 The typical iterative mechanism and the high homology of bimodular cyclodepsipeptide
720 synthetases were also exploited to assemble hybrid NRPSs. In the case of closely related synthetases,
721 inter-module swapping is one of the most effective approaches for altering the incorporated
722 monomer, and thus generating a novel putatively bioactive compound (Brown et al., 2018). For this
723 purpose, the first module of a synthetase is combined with the second module of another system in
724 order to generate new non-natural nonribosomal peptides. Studies reporting this inter-module
725 swapping between cyclodepsipeptide synthetases have employed the above described Tet-on
726 expression system in *A. niger* (Steiniger et al., 2017; Zobel et al., 2016). Module 1 of PSYN NRPS
727 (from *Rosellinia abscondita*), activating either $_{D}$-Lac or $_{D}$-Phe, was fused to module 2 of EnSYN
728 (from *F. oxysporum*) and BeSYN (from *B. bassiana*) activating $_{L}$-Val and $_{L}$-Phe, respectively. The
729 subsequently transformed *A. niger* colonies were able to produce six novel compounds, but with
730 lower yields than for native cyclodepsipeptides. The lower yields observed can also be related to a
731 bottleneck in substrates for different derivatives (Zobel et al., 2016). Using a similar approach, after
732 screening hybrid NRPS resulting from several combinations of modules from different
733 cyclodepsipeptide NRPS (i.e., EnSYN, BeSYN, and BaSYN) in *E. coli*, the most promising
734 combinations were expressed in *A. niger* (Steiniger et al., 2017). The combinations EnSYN-
735 BaSYN.T.C_3, BeSYN-BaSYN.T.C_3, BaSYN-EnSYN.T.C_3, and BaSYN-BeSYN.T.C_3 were
736 successfully expressed, leading to high titers of hybrid cyclodepsipeptides: octa-enniatin B (4 $mg.l^{-1}$),
737 octa-beauvericin (10.8 $mg.l^{-1}$), and hexa-bassianolide (1.3 $g.l^{-1}$). Module and domain swapping
738 between different PKS/NRPS is an innovative approach to generate new compounds, difficult to
739 obtain by chemical synthesis, and displaying interesting biological activity (Süssmuth and Mainz,
740 2017). With this approach, 6 novel compounds were also produced in *Fusarium heterosporum*, after
741 57 swapping combinations with 34 modules from various fungal PKS-NRPS (Kakule et al., 2014).
742 Other notable examples of chimeric compounds include: niduchimaeralin A/B in *A. nidulans* by
743 swapping the PKS-NRPS CcsA from *Aspergillus clavatus* with the Syn2 from *Magnaporthe oryzae*
744 (Nielsen et al., 2016), and predesmethylbassianin A / pretenellin A in *A. oryzae* by swapping between
745 the PKS-NRPS DMBS and TENS from *B. bassiana* (Fisch et al., 2011).

746 Likewise, hybrid PKS-NRPS genes have also been heterologously expressed in filamentous
747 fungi to produce compounds of interest. For example, tennelin was produced in *A. oryzae* after

748 expression of *tenS* PKS-NRPS from *B. bassiana* under the control of the starch-inducible *amyB*
749 promoter (P_{amyB}), leading to up to 5-fold higher yields than in the native fungus (Halo et al., 2008;
750 Heneghan et al., 2010). Under the control of the same promoter, a PKS-NRPS from *Chaetomium*
751 *globosum* was expressed in *A. nidulans*, enabling the production of the potential anti-HIV Sch
752 210972 (Sato et al., 2015). Once again, P_{amyB} was employed to express the 12.4-kb *ACE1* PKS-
753 NRPS from *M. oryzae* in *A. oryzae*, leading to magnaporthepyrone production (Song et al., 2015).
754 Similarly, the aspyridone biosynthetic pathway (*adpA* gene cluster) was reconstituted in *A. oryzae* by
755 cloning the genes of *AdpA* cluster from *A. nidulans* (Wasil et al., 2013). Additionally, two successive
756 PKS and NRPS genes (*glpks3-glnrps7*) in one cluster of *Glarea lozoyensis* were also successfully
757 expressed as a polycistronic mRNA in *A. nidulans* to produce xelolozoyenone (Yue et al., 2015). The
758 biosynthetic pathways of astechrome and macrolactone valactamide were reconstituted also in *A.*
759 *nidulans* through fungal artificial chromosomes.

760 **3.4 Key factors to scale up heterologous production of NRPs**

761 Multiple key factors have to be considered to succeed in NRP heterologous production and to
762 increase the productivity. Each step, from the cloning and the transcription of the long heterologous
763 gene to the functional synthetase, can be a barrier to the production (**Fig. 5**). Some of these limiting
764 steps, and the possible solutions to overcome them, will be discussed in this section.

765 **3.4.1 Incompatibility of the host with foreign NRPS nucleic and amino acid sequences**

766 The incompatibility of the foreign nucleic acid sequences with the host machinery is a major
767 cause of failure of heterologous production. This incompatibility is more pronounced when the native
768 producer is taxonomically distant from the host organism. It can mainly be related to a difference in
769 codon usage slowing down or ending the mRNA translation. As mentioned above, codons of the
770 foreign sequence can be optimized by replacing rare codons by more common ones for the host
771 (Ongley et al., 2013). This approach also enabled, in *Aspergillus* species, a stabilization of the
772 transcripts of heterologous genes by preventing premature polyadenylation and subsequent
773 degradation of mRNA (Tanaka et al., 2014). However, this methodology is time consuming in the
774 design of the optimized sequence and relies on still expensive DNA synthesis technology.

775 At the protein level, the large size of NRPSs increases the possibility to display recognition
776 sequences for host proteases. In particular, the amino acid sequences of inter-module sequence
777 regions, which are largely unfolded (i.e., limited number of α-helix and β-sheet) and display a high

778 degree of flexibility, are therefore more vulnerable to protease activity (Samel et al., 2007; Tarry et
779 al., 2017). Yeasts have been described to express multiple vacuolar proteases, such as
780 endoproteinases A and B, potentially released during cell lysis. To prevent the degradation of the
781 newly-synthesized NRPS, a *S. cerevisiae* strain has been mutated in *pep4* and *prb1* genes encoding
782 the previously mentioned proteases (Jones, 1991; Kealey et al., 1998). The resulting BJ5464 strain
783 has been widely used for heterologous production of various NRPs (Gao et al., 2013, 2012; Haynes
784 et al., 2013; Yu et al., 2013).

785 In addition, non-native large multi-modular complexes may undergo a complete or partial mis-
786 folding, leading to a loss of enzymatic activity due to rapid degradation of the complex or an
787 inappropriate tertiary structure of catalytic sites. This can be solved or attenuated by co-expressing
788 foreign chaperone encoding genes, which are likely to be involved in the folding of the synthetase in
789 the native host (Ongley et al., 2013). Alternatively, it was reported that a lower growth temperature
790 could promote the folding of heterologously produced proteins. For example, an increase in the
791 stability of pcbAB synthetase was demonstrated in *H. polymorpha* grown at 25°C instead of 37°C
792 (Gidijala et al., 2008). Nonetheless, this effect was not observed in *S. cerevisiae* with the same NRPS
793 (Siewers et al., 2009). Moreover, in the enniatin heterologous producer *A. niger*, the effect of the
794 temperature was also investigated among other parameters. It appears that reducing the temperature
795 from 30°C to 26°C did not significantly improve the productivity, especially compared to the impact
796 of the culture medium composition (Richter et al., 2014).

797 **3.4.2 Required enzymes for the biosynthesis of active nonribosomal peptides**

798 In most cases, obtaining the properly folded synthetase is not sufficient to enable the complex
799 to synthesize the compound of interest. Indeed, to be active, NRPSs have to be converted from their
800 *apo* into their *holo* forms through the addition of a flexible phosphopantetheine arm on each
801 thiolation domains by a phosphopantetheinyl transferase. This cofactor is crucial because it carries
802 the monomers previously specifically recognized by the different A-domains. In contrast to
803 filamentous fungi, yeasts (except *S. pombe*) lack a functional PPTase (Schwecke et al., 2006). An
804 endogenous *lys5* PPTase gene, involved in lysine biosynthesis, is expressed in several yeasts but the
805 corresponding enzyme was demonstrated to be unadapted for NRPS activation (Gidijala et al., 2009;
806 Mootz et al., 2002). To complement the lack of a functional PPTase, *sfp* from *B. subtilis* or *npgA*
807 from *A. nidulans*, are generally co-expressed with NRPS genes. NpgA and Sfp have been reported to
808 display a very broad specificity on various NRPSs in diverse fungal and bacterial organisms (Ishiuchi

et al., 2012; Mofid et al., 2002; Quadri et al., 1998; Tobias et al., 2016). The efficiency of these two enzymes was demonstrated in *S. cerevisiae*, and the same order of magnitude of NRP titers was obtained with both of them (Siewers et al., 2009). Conversely, in *E. coli*, the productivity in terms of cyclodepsipeptides could be increased by 4.8- and 14-fold with *npg*A gene instead of *sfp* and endogenous PPTase, respectively (Steiniger et al., 2017). The *npg*A PPTase gene has even been integrated in the genome of a *S. cerevisiae* strain to generate a cellular platform ready to produce NRP compounds (Gao et al., 2013, 2012; Haynes et al., 2013; Yu et al., 2013).

NRPs synthesized by functional synthetases are not necessarily active compounds, and are often only early intermediates of the targeted bioactive product. Therefore, to obtain the final functional product, addition of gene(s) encoding the tailoring enzyme(s) responsible for NRP modification step(s) is required. This is the case for example, for ACV, which is sequentially modified by three tailoring enzymes to give isopenicillin G (**Fig. 6 -B**). To heterologously produce isopenicillin G in yeasts, and not only ACV, addition of the isopenicillin N synthase, isopenicillin N acyl transferase and phenylacetyl CoA ligase genes was required (Awan et al., 2017; Gidijala et al., 2009, 2008). Moreover, the two last above mentioned tailoring enzymes are located in the peroxisome in the native producer *Penicillium chrysogenum*. This parameter can be crucial, as the efficiency of these enzymes is impacted by their location and environment. An absence of compartmentalization of these enzymes in peroxisomes decreased the productivity by more than 60% in heterologous hosts (Gidijala et al., 2009). This trend was confirmed later in *S. cerevisiae* with an increased productivity gained by tagging these enzymes with PTS1 peroxisome addressing signal (Awan et al., 2017).

Gliotoxin produced by *A. fumigatus*, is also a bioactive product resulting from multiple modifications of an NRP diketopiperazine encoded by *gliP* synthetase gene. The first step of the biosynthetic pathway from the NRP to gliotoxin was elucidated through a heterologous expression in *A. nidulans* (Chang et al., 2013; Gardiner and Howlett, 2005). Two tailoring enzymes candidates, namely *gliC* and *gliF* were co-expressed separately with the *gliP* synthetase gene. The disappearance of the precursor NRP in favour of an L-Phe hydroxylated derivative, was only observed in the strain harbouring the cytochrome P450 monooxygenase encoded by *gliC*. This result demonstrated the role of this tailoring enzyme in the first step of gliotoxin biosynthesis.

Similarly, brevianamide F nonribosomal compound, assembled by FtmA (or FtmS) in *A. fumigatus* and *N. fischeri*, is a platform molecule, precursor to multiple compounds. Wunsch and co-

workers have co-transformed *ftmS* in *A. nidulans* with three different prenyltransferase genes from the two previously mentioned fungi. The resulting strains were able to produce, in addition to brevianamide F, up to three supplementary derivative compounds, such as deoxybrevianamide E (Wunsch et al., 2015). With co-expression of *ftmA* with *ftmB*, *ftmC*, *ftmD*, *ftmE*, *ftmF* tailoring enzyme-encoding genes, production of derivatives of brevianamide F was also observed. The NRPS-encoding gene and five tailoring enzyme genes were assembled successively in the same single plasmid, through multiple homologous recombinations in *S. cerevisiae*. All ORFs were positioned in the plasmid under the control of distinct promoters and terminators. After transformation in both *S. cerevisiae* and *A. nidulans*, the enzymatic activity of tailoring enzymes enabled the synthesis of multiple precursors of spyrotriprostatin starting from brevianamide F: tryprostatin B, tryprostatin A, and fumitremorgicin (Tsunematsu et al., 2013a).

3.4.3 How to address NRP precursors deficiency?

In the classical ribosomal pathway, only 20 proteinogenic α-amino acid residues are incorporated into proteins. In addition to the proteinogenic α amino acid residues, up to 500 distinct monomers were reported to constitute NRPs building blocks (Caboche et al., 2010). Among this tremendous biodiversity of monomers, some are directly recognized and incorporated as such, while others are modified during the peptide synthesis by the activity of a NRPS domain. This incredible variety of precursor monomers is the root of the wide diversity of structure and activity of NRPs (Challis and Naismith, 2004; Schauwecker, 2003). On the other hand, this feature raises a serious issue regarding heterologous production, as hosts lack the ability to synthesize most of these non-common monomers. A possible approach to complement monomers deficiency relies on the cloning of the genes involved in the monomer biosynthesis pathway. The monomer $_D$-hydroxyisovaleric acid ($_D$-Hiv) is for instance not synthesized by *A. niger* and poorly produced by *S. cerevisiae*. $_D$-Hiv is a precursor engaged in the biosynthesis of cyclodepsipeptides, such as enniatin, beauvericin and bassianolide. It can be obtained by the modification of α-ketovaleric acid by α-ketoisovalerate reductase encoded by *kivR* gene in *B. bassiana* or *F. oxysporum*, among others (Richter et al., 2014; Yu et al., 2013). A two-fold increase of productivity was obtained with the expression of *kivR* gene in a *S. cerevisiae* strain heterologously producing beauvericin (61.7 and 33.8 $mg.l^{-1}$ with or without *kivR*, respectively) (Yu et al., 2013). In *A. niger*, the enniatin heterologous production has also been made possible with the co-expression of the missing *kivR* encoding α-ketoisovalerate reductase with *esyn1* NRPS gene (Richter et al., 2014).

On the other hand, a precursor deficiency can also be solved by feeding the corresponding compound directly into the culture medium. This approach avoids time consuming gene cloning steps but relies on the host strain being able to take up the precursor from the medium. The complementation of the above mentioned *A. niger*, expressing foreign enniatin synthetase, was also achieved by feeding it with D-Hiv instead of co-expressing *kivR*. The feeding experiment was quite successful, as it yielded a high enniatin B titer of 950 $mg.l^{-1}$ after optimization of feeding conditions and culture medium composition. Under the same conditions, the strain fed with 20 mM D-Hiv was produced 25% more than the strain transformed with *kivR* (Richter et al., 2014). Nonetheless, the cloning approach remains more cost effective due to the high cost of D-Hiv (Yu et al., 2013). Moreover, some adenylation domains display a broader specificity that can be leveraged to generate alternative final products by feeding with alternative precursors. For instance, enniatin B is preferably produced by feeding L-valine in the medium, whereas the alternative enniatin A production is enhanced by feeding with L-leucine or L-isoleucine (Richter et al., 2014). In another filamentous fungus, *A. nidulans*, asperphenamate production was knocked down by mutating the *apmA* gene responsible for N-benzoylphenylalaninol biosynthesis. This production was subsequently restored by feeding the same strain with the deficient precursor (Li et al., 2018). In yeast, the relevance of precursor supplementation in the culture medium has been highlighted in *H. polymorpha* expressing *pcbAB* NRPS gene. The production of ACV was not possible due to the incapacity of *H. polymorpha* to synthesize the precursor aminoadipic acid. Supplementation of the culture medium with this monomer enabled to unlock the production of ACV (Gidijala et al., 2009). Nonetheless, the use of a precursor feeding strategy is not always a guarantee of success as some negative feedback mechanisms can occur and hinder the compound production. For instance, the feeding of the NRP-precursor 2-aminobutyric acid in the culture medium of C. asteris, was recently demonstrated to negatively impact the production of the astin C compound as well as the development of the fungus (Vassaux et al., 2019).

Furthermore, the highly developed secondary metabolism of most of the filamentous fungi can hogged available nutrients and precursors. This can lead to a decrease in availability of some monomers and even become a limiting factor for the production of heterologous NRPs. The modification of the metabolic fluxes and the deletion of SM biosynthetic pathways consuming large amount of cell nutrients and energy can be favourable for the production of the foreign compound of interest (Luo et al., 2016). Moreover, reducing the secondary metabolite metabolisms of the selected host can facilitate the detection and the purification of the targeted compound by decreasing the

presence of unwanted products (He et al., 2018). Multiple methodologies have been developed to knock out entire SM gene clusters in *A. nidulans*. Among the 71 BGCs predicted in *A. nidulans* (Inglis et al., 2013), the clusters related to the biosynthesis of two major native metabolites of this fungi, the emericellamide and the sterigmatocystin, have been deleted to generate a "cleaner" expression host for PKS/NRPS heterologous production (Chiang et al., 2013).

3.4.4 Heterologous NRP toxicity mitigation

A further challenge facing NRP heterologous production is related to the toxicity of some compounds towards the selected host. Native producers have developed various strategies to prevent this toxicity, such as specific secretion mechanisms, compartmentalization of the toxic products, or expression of genes encoding enzymes able to induce a resistance (Bond et al., 2016; Chanda et al., 2009; Martín et al., 2005). These mechanisms are often specific to the toxic NRP product and therefore may be only present in the native producer and not in the chosen host (Nielsen and Nielsen, 2017). The most straightforward strategy to overcome this issue consists to express the NRPS gene under the control of an inducible promoter. This approach enables the choice of the period of production of the toxic compound and to dissociate it from the cell growth period. Hence, GAL1-10, amyB and Tet-on inducible promoters were used to express NRPS genes, respectively in *S. cerevisiae* (Siewers et al., 2009), *A. nidulans* (Fujii et al., 1996), and *A. niger* (Richter et al., 2014).

Another common approach consists of transferring into the host not only the NRP biosynthesis related genes, but also the specific resistance genes pre-identified into the genome of the native producer. The mechanisms of self-resistance to a produced toxic compound can be classified into three categories of mode-of-operation: (1) by inducing the compound efflux, (2) through enzymatic deactivation of the toxic product, or (3) with a targeted alteration of the drug receptors (Cundliffe and Demain, 2010). Considering that the latter are quite uncommon, and that the deactivation of the toxic product can also impair the targeted activity of the compound, the implementation of an efflux system is often the most straightforward approach to confer a resistance (Keller, 2015). NRPs are likely to be small structures and accordingly possibly naturally secreted by some organisms. Nonetheless, numerous studies reported a poor secretion of NRPs by heterologous strains such as *H. polymorpha* (Gidijala et al., 2009), *S. cerevisiae* (Awan et al., 2017), and *A. niger* (Zobel et al., 2016). Therefore, expression of a foreign NRP specific transport protein can be required to improve the secretion procedure (do Valle Matta et al., 2001). However, the identification of resistance gene (regardless of their mode-of-operation) in the native producer is not always easy. In filamentous

fungi, genes conferring resistance to toxic secondary metabolites have been highlighted multiple time within the corresponding gene cluster. For instance, in the gliotoxin BGC of *A fumigatus*, the *gliA* and *gliT* genes respectively encode an efflux pump and a gliotoxin oxidoreductase which can reduce the toxicity of the compound (Dolan et al., 2015; Sharma and Chowdhary, 2017). Nonetheless, these resistance genes are not systematically located within the corresponding gene cluster, and the annotation methods are often far to be sufficiently efficient to identify precisely resistance genes, especially in fungal organisms (Tran et al., 2019).

There are further options to protect the heterologous host from the toxicity of the foreign compound, such as extracting the product from the culture medium. The toxic compound can for instance be extracted by introducing a layer of organic solvent on the top of the culture, where the product can diffuse (Bond et al., 2016; Rodriguez et al., 2014). Some absorber resins are also able to bind the toxic compound when they are added to the culture broth. This was for instance demonstrated with the Diaion HP20 resin binding the mycotoxin cercosporamide contained in a fungal fermentation broth, thus enabling to increase by 100-fold the final product yield (Singh et al., 2010).

3.4.5 From the heterologous host to the *in vitro* production assays: a viable approach?

3.4.5.1 Approach and methodologies

The *in vitro* NRP synthesis constitutes a cell-free production system where the enzymes involved in the biosynthetic pathway (NRPS gene, tailoring enzymes) are purified and subsequently incubated with the precursor monomers to generate the desired compound. The *in vitro* approach is a suitable solution to address most of the issues described above. Indeed, it addresses the deficiency in NRP precursors, but also circumvents issues of NRP toxicity towards the host strain. In order to perform *in vitro* assays, the synthetase and/or associated tailoring enzymes need to be purified in soluble form and in sufficient quantity from the heterologous host. Addition of a tag sequence, either on the C-terminal or on the N-terminal part of the heterologously expressed synthetase, is highly recommended to facilitate further extraction-purification steps. The use of 6xHis tag (Brandenburger et al., 2017; Gao et al., 2013; Haynes et al., 2013) or Flag tag (Gao et al., 2012), respectively purified through metal or antibody affinity chromatography, have been reported in previous *in vitro* NRPS studies. Alternatively, NRPSs were also purified by size exclusion chromatography as their high molecular weights facilitates their separation from other host proteins (Fickers et al., 2008; Lawen and Zocher, 1990).

Thereafter, it is possible to assess the presence of the phosphopantetheine arm on thiolation domains through three approaches. The most recently developed method consists to generate a peptide mixture by a trypsin digestion of the proteins contained in the cell extract, before to separate fractions from the peptide mixture through a strong cation exchange chromatography. Subsequently, a MS/MS analysis enables to detect into the different fractions the presence of unique marker ions derived from the NRPS-bonded phosphopantetheine arm (Bumpus et al., 2009). Alternatively, the presence of the phosphopantetheine arm on the *holo*-NRPS can also be evaluate through the thioester-binding assay measuring the consumption of radiolabeled [^{14}C] amino-acids incorporated in the synthesized NRP (Schwarzer et al., 2001). An older technic consists to perform an alkaline phosphatase treatment to hydrolyze the phosphopantetheine part from the *holo*-NRPS. This treatment releases a pantothenic acid which can be quantified through a microbiological assay against *Lactobacillus plantarum* (Lawen and Zocher, 1990). Regarding to A-domains, the most common approach to confirm their enzymatic activity consists to quantify the incorporation of ^{32}PPi into ATP in presence of the precursor amino acids (ATP pyro-phosphate exchange radioassay) (Brandenburger et al., 2017; Fickers et al., 2008). Alternatively, a nonradioactive method has also been developed enabling to evaluate the A-domains activity through a MS/MS analysis. This technic relies on the measurement of the isotopic back exchange of the pyrophosphate into a [^{18}O]-labeled ATP (Phelan et al., 2009). Recently, Duckworth and coworkers (2016) have developed another method based on the measurement of the pyrophosphate released during the adenylation reaction. The pyrophosphate released subsequently to the action of an active A-domain is first cleaved into an inorganic phosphate by an inorganic pyrophosphate. Subsequently, a purine nucleoside phosphorylase uses this inorganic pyrophosphate in a reaction converting 7-methylthioguanosine into 7-methylthioguanine which can be monitored with a spectrometric-based method at 360 nm (Duckworth et al., 2016). Finally, the functionality of the *holo*-NRPS complex can be confirmed with the identification of the final NRP by using appropriate analytical method (LC-MS and/or NMR).

3.4.5.2 Successful *in vitro* assays for reconstitution/production of fungal NRPs

Genome mining and *in silico* approaches have enabled, in *C. subvermispora*, to assign the basidioferrin siderophore synthesis to the *CsNPS2* NRPS-encoding gene. Experimental validation of these predictions has been performed through heterologous production coupled with *in vitro* assays. *CsNPS2* was expressed in *A. niger* and subsequently purified through metal affinity chromatography. The single A-domain of CsNPS2 displayed specificity for two substrates: N_5-acetyl-N_5-hydroxy-$_L$-ornithine ($_L$-AHO) and N_5- cis-anhydromevalonyl-N_5-hydroxy-$_L$-ornithine ($_L$-AMHO). Nonetheless,

997 basidioferrin was only detected after 24h of *in vitro* assay with 1 mM L-AHO and not with L-AMHO.
998 These results confirmed the role of *CsNPS2* in the basidioferrin biosynthesis (Brandenburger et al.,
999 2017).

1000 Several NRPSs were also heterologously expressed and purified from *S. cerevisiae* for the
1001 purpose of *in vitro* assays: AnaPS (Gao et al., 2012), TqaA (Gao et al., 2012), AspA (Gao et al.,
1002 2013), ArdA (Haynes et al., 2013). The presence of the *npgA* gene in the genome of the yeast was
1003 compulsory to produce the PPTase enabling to convert the synthetase from the *apo* to the *holo*
1004 activated form. *In vitro* assays with 10 µM pure AspA with its monomer Ant (1 mM) and L-Trp (1
1005 mM) generated a tripeptide, the iterative mode-of-operation of this bimodular NRPS (Gao et al.,
1006 2013). The TqaA synthetase is to date the largest NRPS protein (450 kDa) whose activity has been
1007 reconstituted *in vitro* in its active form (Gao et al., 2012). The ArdA synthetase from *N. fischeri*,
1008 displaying high homology with TqaA, was also used to perform *in vitro* assays after being purified
1009 from a heterologous *S. cerevisiae* producer. This 430-kDa bimodular NRPS was demonstrated *in*
1010 *vitro* to sufficiently assemble Ant, L-Ala, and L-Trp into ardeemin FQ (Haynes et al., 2013).

1011 In another *in vitro* study aiming at obtaining NRP derivatives, Tsunematsu et al. (2013) first
1012 produced and purified the different intermediates of the spirotryprostatin biosynthetic pathways in a
1013 heterologous *S. cerevisiae*. In the meantime, tailoring enzymes FtmG and FqzB were also purified
1014 from heterologous yeast. With the purpose of deciphering the role of these tailoring enzymes in the
1015 final step of spirotryprostation synthesis, *in vitro* assays were performed with the purified precursors
1016 mentioned above. As a result of this work, *ftmG* encoding a cytochrome P450 has been shown to be
1017 involved in spiro-ring formation in spirotryprostatin B, while the FAD-dependent monooxygenase
1018 encoded by *fqzB* was shown to be responsible for the spiro-carbon formation in spirotryprostatin A
1019 (Tsunematsu et al., 2013a).

1020 The *in vitro* NRP assembling strategy is especially useful as a complement to heterologous
1021 production by facilitating the elucidation of synthetase mode-of-operation or tailoring enzymes
1022 function. Nonetheless, due to the very low amounts of NRPS obtained through affinity
1023 chromatography purification, this approach is not promising for increasing NRP yields. Moreover,
1024 the cost incurred by the addition of precursor monomers and ATP largely prevents its applications for
1025 large scale NRP production.

1026 **4. Penicillin heterologous production: a NRP model system**

Due to the clinical and commercial importance of β-lactam antibiotics, researchers have attempted for years to develop their heterologous production in hosts renowned for being readily cultivate at industrial scale, and easy to handle through genetic engineering. This multi-decade work has positioned the NRP-derived penicillin compounds as a model system regarding NRP heterologous production. These efforts have enabled both to better understand the ins and outs of the penicillin biosynthetic pathway and to develop various strategies to implement heterologous production in fungal cell factories.

Penicillin β-lactam antibiotics are naturally produced by some *Penicillium* species, especially *P. chrysogenum* and to a lesser extent by some *Aspergillus* species such as *A. nidulans* (R W Newbert et al., 1997; Suárez and Peñalva, 1996). In *P. chrysogenum*, the elucidated penicillin biosynthetic pathway relies on the joint action of several enzymes, whose corresponding genes are mainly located within the same gene cluster (**Fig. 6 -A**). The *pcbAB* gene encodes a 3-module NRPS that is responsible for assembling ACV (**Fig. 6 -B**). This nonribosomal peptide precursor of penicillin, is then modified into isopenicillin (IPN) by an isopenicillin N synthase. This intermediate is further converted by an isopenicillin N acyl transferase (IAT) and a phenylacetyl CoA ligase (PCL) into penicillin G (Gidijala et al., 2009). The penicillin V in which the benzyl group is replaced by a phenoxymethyl group is a commercial analog of the penicillin G displaying an enhanced stability in acidic environment. This analog is preferably produced by *P. chrysogenum* when phenoxyacetic acid is added in the fermentation medium (Ball et al., 1978). This appears reasonable because the production of low amounts of penicillin compounds, and the presence of enzymes sharing a high level of sequence identity with those from *P. chrysogenum*, had been reported in (Suárez and Peñalva, 1996). An associated positive impact of these native enzymes cannot be ruled out to explain the success of this heterologous production.

The first transfer of the penicillin biosynthetic pathway in a heterologous host was achieved into *N. crassa* and *A. niger*. The cosmid pCX3.2 containing the ACVS encoding gene cluster was cloned into both heterologous hosts, leading directly to production of penicillin. The presence of the produced antibiotic was successfully assessed through a bioactivity assay against *B. subtilis*. *A. niger* was reported to produce much more penicillin V than *N. crassa* (up to 2.3 $\mu g.ml^{-1}$ versus 0.170 $\mu g.ml^{-1}$) (Smith et al., 1990). One explanation could be the lack of ACV precursor availability in *N. crassa*. A second hypothesis rests on the fact that *A. niger* and *P. chrysogenum* are closely related species both belonging to the order Eurotiales, unlike *N. crassa* which is associated to Sordariales.

A few years later, the penicillin biosynthetic pathway was also reconstituted into the methylotrophic yeast *H. polymorpha*, which is well-known for its superior fermentation properties even at industrial scale (Stöckmann et al., 2009). The 12-kb *pcbAB* NRPS gene was inserted, under the control of the strong methanol-inducible promoter from alcohol oxidase gene (P_{AOX}) in pHIPZ4 plasmid. *H. polymorpha* was transformed with the linearized plasmid in order to integrate the gene at the AOX locus. Subsequently, heterologous expression of ACV synthetase (ACVS) was assessed directly at the protein level by Western blot with antibodies directed to ACVS amino-acid sequence. As described in the section 3, the homologous PPTase Lys5 of the methylotrophic yeast was not adapted to activate ACVS, and an additional *sfp* PPTase encoding-gene was inserted. Subsequently, IPNS, IAT and PCL-encoding genes from *P. chrysogenum* were integrated into the genome step by step. Penicillin G productivity of the resulting yeast was similar to those obtained in the native producer NRRL1951 (1.1 and 1.0 $\mu g.ml^{-1}$, respectively). In the same yeast, deletion of the native *PEX3* gene, responsible for the peroxisome proliferation, led to a 64% reduced yield (0.4 $\mu g.ml^{-1}$) (Gidijala et al., 2009). The importance of the peroxisomal location of IAT and PCL had previously been highlighted in *P. chrysogenum* and *A. nidulans* (Spröte et al., 2009).

In subsequent work, *S. cerevisiae* was also selected to heterologously produce ACV (Siewers et al., 2009). *pcbAB* was inserted into a replicative multi-copy pESC by TAR cloning under the control of the galactose inducible promoter GAL1/GAL10 ($P_{GAL1/GAL10}$). The resulting construct bearing an *URA3* selection gene was used to transform a *S. cerevisiae* Δ*ura3* strain. One of the assets of pESC is the presence of the bi-directional $P_{GAL1/GAL10}$ enabling the cloning of a second gene. This allows the cloning of the PPTase encoding gene in the same vector as the NRPS gene (if the NRPS is encoded by a single ORF). Under galactose induction, the transformed strain carrying this plasmid was able to express functional ACVS (transcription assessed by qPCR and detection of the NRP by LC-MS analysis). A codon optimization of the 5' portion of the ACVS gene was also performed leading to a slight, but not significant, increase of the productivity. On the other hand, *pcbAB* was also integrated into a long terminal repeats (LTR) retrotransposons locus of *S. cerevisiae*. LTR retrotransposon are transposable elements present in high number and spread all over the yeast genome, thus making these sites suitable for multi-copy genomic integration (Maury et al., 2016). Yields obtained with genomic integration were 6-fold lower than in the multi-copy PESC plasmid, demonstrating the benefits of the latter expression system. A few years later, another ACV producing *S. cerevisiae* was obtained with the cloning of *pcbAB* and *npgA* gene into a low-copy centromeric plasmid (Awan et al., 2017). Multiple pairs of promoters were tested to enhance the expression of both the *pcbAB* and

npgA genes. The combination of *pcbAB* and *npgA*, respectively under the control of P_{TDH3} and P_{PGK1} constitutive promoters, led to a 4-fold higher ACV productivity than with $P_{GAL1/GAL10}$ in PESC plasmid (≈0.28 versus ≈0.07 $\mu g.ml^{-1}$). This demonstrated the importance of promoter choice for each gene to be expressed, and the underestimated requirement for compatibility of each promoter/gene pair. Subsequently, the complete biosynthetic pathway to the penicillin G was reconstituted by addition of tailoring enzymes IPN, IAT and PCL. Expression of the tailoring enzymes were also evaluated under multiple combinations of promoters for the sake of optimizing the production. The resulting strain was able to efficiently produce and secrete penicillin G, but with yields still two fold less than in heterologous *H. polymorpha* (Gidijala et al., 2009).

With the purpose of transferring the multiple genes involved in the penicillin synthesis from *P. chrysogenum* into *A. nidulans* SAA248 lacking a native penicillin gene cluster, Unkles and coworkers have exploited the previously described properties of the viral 2A peptide. By inserting viral 2A peptide between each single ORF, it is possible to express from the same plasmid a polycistronic mRNA under the control of a single strong/inducible promoter (Kim et al., 2011). Viral 2A sequences were inserted between *pcbAB*, IPNS gene (*pcbC*), and IAT gene (*penDE*). The resulting 15.5-kb sequence was cloned by *in vivo* homologous recombination in *S. cerevisiae*, into pJET1.2 plasmid under the control of xylose inducible promoter. The obtained plasmid was then transferred to *A. nidulans*, leading to the transcription of the 15.5-kb polycistronic mRNA. Subsequently, each individual enzyme and especially the massive NRPS were cleaved and their joint action enabled the synthesis of penicillin G. Notably, the 331-T25 mutant having integrated multiple copies of the genes, was able to produce 3-fold more penicillin G than the wild type *A. nidulans* strain 0760 (0.687 versus 0.234 $\mu g.ml^{-1}$) (Unkles et al., 2014). Interestingly, in addition to the heterologous IAT enzyme, a second native IAT was identified in the transformed SAA248 strain. Even if *A. nidulans* SAA248 has lost is ability to synthesize penicillin like other *A. nidulans* strain, the presence of genes encoding enzymes involved in the penicillin biosynthetic pathway seems to persist in its genome (Suárez and Peñalva, 1996). An associated positive impact of these native enzymes cannot be ruled out to explain the success of this heterologous production.

It is not straightforward to compare the penicillin production yields obtained with the different heterologous strains described in this section. Nonetheless, authors have each time compared these yields with those observed either in the *P. chrysogenum* NRRL1951 or *A. nidulans* 0760 native penicillin producers, enabling some cross-study comparisons. Regarding to yeast heterologous systems, *H. polymorpha* was reported to produce similar penicillin G amount to *P. chrysogenum*

NRRL1951 strain (~ 1 µg.ml^{-1}) (Gidijala et al., 2009). In contrast, *S. cerevisiae* was half efficient, probably due to the considerably less developed peroxisomal system compared to the methylotrophic yeast (Awan et al., 2017). Regarding to filamentous fungi heterologous systems, the 331-T25 mutant of *A. nidulans* SAA248 expressing penicillin biosynthetic genes coming from *P. chrysogenum* was 3-fold more efficient than the native penicillin G producer *A. nidulans* 0760 (~ 0.69 µg.ml^{-1}), but still 30% less productive than *P. chrysogenum* NRRL1951 (Unkles et al., 2014). In terms of heterologous penicillin V production, wide variations have been observed between the two described hosts with up to 13-fold higher productivity in *A. niger* compared to *N. crassa*. The highest penicillin V concentrations reported in the heterologous *A. niger* strain was likely due to the close taxonomic relatedness of this strain with *P. chrysogenum* (Smith et al., 1990).

Over many years, the heterologous production of penicillin has been a long-standing test case of study. It enabled researchers to deal with most of the key factors limiting NRP heterologous production and to unlock some of them. Unfortunately, in terms of productivity, the heterologous production approach is decades behind on current industrial penicillin G production systems. Indeed, several studies were conducted to generate and screen thousands of *P. chrysogenum* mutants, with the aim of improving production yields. Industrial production, with these selected strains, has been optimized for years in terms of strain productivity and fermentation process (R. W. Newbert et al., 1997; Vournakis and Elander, 1983). The analysis of the genome of the strains with an improved penicillin production revealed the presence of multiple copies of the penicillin biosynthetic genes, which is likely to contribute in the increased productivity (Fierro et al., 2006; Müller et al., 2018; R W Newbert et al., 1997). Another characteristic of high-penicillin producing strains is the loss of several gene clusters involved in secondary metabolite synthesis which were detected in the wild-type *P. chrysogenum* NRRL195. For instance, the improved penicillin producer *P. chrysogenum* Wisconsin 54-1255 lacks the ability of producing sorbicillin and penitrinic acid which are present in the genome of the parental NRRL195 strain. The deletion of secondary metabolite biosynthetic pathways also contributed to improve the strain productivity for β-lactam antibiotics (Salo et al., 2016, 2015). Although the current penicillin G industrial production yields are not available in the literature, these yields were reported few years ago to be already more than 1000-fold higher than those obtained with the wild-type *P. chrysogenum* NRRL1951 (Harris et al., 2009).

5. Concluding remarks

The recent development of high-throughput sequencing technologies combined with substantial drop in sequencing costs have driven an impressive increase in the number of available fungal genomes. This factor coupled with the improvement of genome annotating methodologies has partially unlocked the SM production potential of fungi. Hence, efficient bioinformatics tools now enable a more accurate analysis in terms of nonribosomal pathway identification and substrate specificity predictions. Nonetheless, the prediction robustness is lowered because already elucidated fungal NRPs and NRPSs remain under-represented in dedicated databases as compared to bacterial ones. Consequently, the predictions based on genomic sequences need to be validated by structural elucidation of the NRPs which implies its efficient production.

The production without genetic manipulation of NRP compounds identified *in silico* is not systematically possible in native hosts due to the existence of cryptic biosynthetic gene clusters and/or incompatibility between the strain and the laboratory cultivation conditions. Therefore, the heterologous production approach is often the most appropriate strategy to deal with this issue. Multiple fungal cellular platforms have been developed to heterologously produce foreign fungal nonribosomal peptides in particular. Among others, the yeast *S. cerevisiae* and filamentous fungi related to *Aspergillus* spp. have been widely used. Heterologous production of NRPs is particularly challenging due to a plethora of potential limiting steps. Combining multiple strategies to optimize each of them can greatly increase the chances of successful production. Especially, the transfer of large NRPS sequences encompassing tens of kb is one of the main hurdles to overcome. Several methodologies, such as TAR cloning, ExRec, Gibson assembly, USER cloning, and overlap extension PCR, have been developed to enable the reconstruction of single large NRPS genes or gene clusters. This laborious and time-consuming engineering of large NRPS biosynthetic pathways can be avoided by *de novo* DNA synthesis, but this remains prohibitively costly in many cases. However, considering that costs of gene synthesis are continuously decreasing, an intensification of the use of this rapid methodology for transferring foreign NRPS genes and gene clusters can be expected in the upcoming years.

The correct transfer of the targeted gene in the heterologous host is not a guarantee of success for the production of the related NRP. For instance, Harvey et al. (2018) have transferred 41 cryptic fungal gene clusters to *S. cerevisiae* but only 54% have been processed by the yeast to lead to a detectable compound. This result, while promising, demonstrates the challenges related to the heterologous expression of large fungal gene clusters. Nonetheless, the efficiency of the heterologous production of NRPs can be greatly improved by taking into account some key factors. First, the

transcription rate can be improved by using an appropriate promoter, and sometimes with adequate initiation factors. Translation and proper folding of the protein are also crucial factors to consider. Optimization of this can be achieved by expression of foreign chaperones enzymes or with different cultivation conditions. It is also important to protect the foreign NRPS from the proteolytic activity of the selected host. Once the properly folded NRPS is obtained and activated via a homologous or heterologous phosphopantetheinyl transferase, the assembly of the NRP can be considered. At this last stage, NRP precursor deficiency or NRP toxicity against the selected hosts can also become limiting factors. The first one can be overcome by feeding or cloning of the missing genes related to precursor biosynthesis, while the second may be prevented by conferring secretion ability to the host with the cloning of transporter-associated genes, for example.

Despite these drawbacks, heterologous production strategies were successfully employed to identify novel NRP compounds, to decipher NRPS modes-of-operation and in some cases, to improve productivity. This approach is especially relevant to producing compounds for which the corresponding genes are cryptic or weakly expressed in native organisms for which no genetic tools are available yet. Conversely, transferring an NRP biosynthetic pathway to a heterologous host is often not sufficient to overcome production yield obtained with a native producer already optimized and well industrially established strain as in the penicillin case. Overall, the NRP production rates currently obtained with heterologous expression do not yet reach levels suitable for industrial production. Chemical synthesis remains in most of the cases, the more competitive way to obtain peptides and especially regarding linear peptides including only proteinogenic amino acids. A possible exception are certain high-value compounds, which are cyclic, or include unusual building block and/or monomers which are further modified by tailoring enzyme. Indeed, these specific characteristics are often difficult to obtain via chemical synthesis and should be obtained more easily through a heterologous expression of the corresponding NRPS-encoding gene.

6. Conflict of Interest

Authors of the manuscript declare there is no conflict of interests.

7. Acknowledgments

This work was supported by the CPER Alibiotech project; the INTERREG V France-Wallonie-Vlaanderen Project SmartBioControl/BioScreen; the Fonds pour la Formation à la Recherche dans l'Industrie et l'Agriculture (FRIA); and the international laboratory Microbial Secondary Metabolites

1214 MOM (University of Lille/Gembloux Agro-Bio Tech University of Liege). Grant PhD co-funding of
1215 the University of Liege (Welcome Grant Philippe Jacques) and the University of Lille.

8. References

Agmon, N., Mitchell, L.A., Cai, Y., Ikushima, S., Chuang, J., Zheng, A., Choi, W.-J., Martin, J.A., Caravelli, K., Stracquadanio, G., Boeke, J.D., 2015. Yeast Golden Gate (yGG) for the Efficient Assembly of S. cerevisiae Transcription Units. ACS Synth. Biol. 4, 853–9. https://doi.org/10.1021/sb500372z

Agüero-Chapin, G., Pérez-Machado, G., Sánchez-Rodríguez, A., Santos, M.M., Antunes, A., 2016. Alignment-Free Methods for the Detection and Specificity Prediction of Adenylation Domains, in: Methods in Molecular Biology (Clifton, N.J.). pp. 253–272. https://doi.org/10.1007/978-1-4939-3375-4_16

Akada, R., Kitagawa, T., Kaneko, S., Toyonaga, D., Ito, S., Kakihara, Y., Hoshida, H., Morimura, S., Kondo, A., Kida, K., 2006. PCR-mediated seamless gene deletion and marker recycling in Saccharomyces cerevisiae. Yeast 23, 399–405. https://doi.org/10.1002/yea.1365

Alberti, F., Foster, G.D., Bailey, A.M., 2017. Natural products from filamentous fungi and production by heterologous expression. Appl. Microbiol. Biotechnol. 101, 493–500. https://doi.org/10.1007/s00253-016-8034-2

Ames, B.D., Liu, X., Walsh, C.T., 2010. Enzymatic processing of fumiquinazoline F: A tandem oxidative-acylation strategy for the generation of multicyclic scaffolds in fungal indole alkaloid biosynthesis. Biochemistry 49, 8564–8576. https://doi.org/10.1021/bi1012029

Antipov, D., Korobeynikov, A., McLean, J.S., Pevzner, P.A., 2016. HybridSPAdes: An algorithm for hybrid assembly of short and long reads. Bioinformatics 32, 1009–1015. https://doi.org/10.1093/bioinformatics/btv688

Anyaogu, D.C., Mortensen, U.H., 2015. Heterologous production of fungal secondary metabolites in Aspergilli. Front. Microbiol. 6, 1–6. https://doi.org/10.3389/fmicb.2015.00077

Araujo, R., Sampaio-Maia, B., 2018. Fungal Genomes and Genotyping, in: Advances in Applied Microbiology. pp. 37–81. https://doi.org/10.1016/bs.aambs.2017.10.003

Awan, A.R., Blount, B.A., Bell, D.J., Shaw, W.M., Ho, J.C.H., McKiernan, R.M., Ellis, T., 2017. Biosynthesis of the antibiotic nonribosomal peptide penicillin in baker's yeast. Nat. Commun. 8, 1–8. https://doi.org/10.1038/ncomms15202

Ball, A.P., Gray, J.A., Murdoch, J.M., Ball, A.P., Gray, J.A., Murdoch, J.M., 1978. The Natural Penicillins — Benzylpenicillin (Penicillin G) and Phenoxymethylpenicillin (Penicillin V), in: Antibacterial Drugs Today. Springer Netherlands, pp. 6–18. https://doi.org/10.1007/978-94-011-8004-7_3

Barajas, J.F., Blake-Hedges, J.M., Bailey, C.B., Curran, S., Keasling, J.D., 2017. Engineered polyketides: Synergy between protein and host level engineering. Synth. Syst. Biotechnol. 2, 147–166. https://doi.org/10.1016/j.synbio.2017.08.005

Bashir, A., Klammer, A.A., Robins, W.P., Chin, C.S., Webster, D., Paxinos, E., Hsu, D., Ashby, M., Wang, S., Peluso, P., Sebra, R., Sorenson, J., Bullard, J., Yen, J., Valdovino, M., Mollova, E., Luong, K., Lin, S., Lamay, B., Joshi, A., Rowe, L., Frace, M., Tarr, C.L., Turnsek, M., Davis, B.M., Kasarskis, A., Mekalanos, J.J., Waldor, M.K., Schadt, E.E., 2012. A hybrid approach for the automated finishing of bacterial genomes. Nat. Biotechnol. https://doi.org/10.1038/nbt.2288

Beites, T., Mendes, M. V., 2015. Chassis optimization as a cornerstone for the application of synthetic biology based strategies in microbial secondary metabolism. Front. Microbiol. 6, 1–10.

https://doi.org/10.3389/fmicb.2015.00906

Beld, J., Sonnenschein, E.C., Vickery, C.R., Noel, J.P., Burkart, M.D., 2014. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. Nat. Prod. Rep. 31, 61–108. https://doi.org/10.1039/c3np70054b

Belshaw, P.J., Walsh, C.T., Stachelhaus, T., 1999. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. Science (80-.). https://doi.org/10.1126/science.284.5413.486

Blin, K., Kim, H.U., Medema, M.H., Weber, T., 2017a. Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. Brief. Bioinform. https://doi.org/10.1093/bib/bbx146

Blin, K., Medema, M.H., Kottmann, R., Lee, S.Y., Weber, T., 2017b. The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. Nucleic Acids Res. 45, D555–D559. https://doi.org/10.1093/nar/gkw960

Blin, K., Wolf, T., Chevrette, M.G., Lu, X., Schwalen, C.J., Kautsar, S.A., Suarez Duran, H.G., De Los Santos, E.L.C., Kim, H.U., Nave, M., Dickschat, J.S., Mitchell, D.A., Shelest, E., Breitling, R., Takano, E., Lee, S.Y., Weber, T., Medema, M.H., 2017c. AntiSMASH 4.0 - improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res. 45, W36–W41. https://doi.org/10.1093/nar/gkx319

Bode, H.B., Brachmann, A.O., Jadhav, K.B., Seyfarth, L., Dauth, C., Fuchs, S.W., Kaiser, M., Waterfield, N.R., Sack, H., Heinemann, S.H., Arndt, H.D., 2015. Structure Elucidation and Activity of KolossinA, the D -/ L -Pentadecapeptide Product of a Giant Nonribosomal Peptide Synthetase. Angew. Chemie - Int. Ed. 54, 10352–10355. https://doi.org/10.1002/anie.201502835

Boecker, S., Grätz, S., Kerwat, D., Adam, L., Schirmer, D., Richter, L., Schütze, T., Petras, D., Süssmuth, R.D., Meyer, V., 2018. Aspergillus niger is a superior expression host for the production of bioactive fungal cyclodepsipeptides. Fungal Biol. Biotechnol. 5, 4. https://doi.org/10.1186/s40694-018-0048-3

Boeke, J.D., Trueheart, J., Natsoulis, G., Fink, G.R., 1987. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. Methods Enzymol. 154, 164–75.

Bok, J.W., Ye, R., Clevenger, K.D., Mead, D., Wagner, M., Krerowicz, A., Albright, J.C., Goering, A.W., Thomas, P.M., Kelleher, N.L., Keller, N.P., Wu, C.C., 2015. Fungal artificial chromosomes for mining of the fungal secondary metabolome. BMC Genomics 16, 1–10. https://doi.org/10.1186/s12864-015-1561-x

Bond, C., Tang, Y., Li, L., 2016. Saccharomyces cerevisiae as a tool for mining, studying and engineering fungal polyketide synthases. Fungal Genet. Biol. 89, 52–61. https://doi.org/10.1016/j.fgb.2016.01.005

Bozhüyük, K.A.J., Linck, A., Tietze, A., Kranz, J., Wesche, F., Nowak, S., Fleischhacker, F., Shi, Y.N., Grün, P., Bode, H.B., 2019. Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. Nat. Chem. 11, 653–661. https://doi.org/10.1038/s41557-019-0276-z

Brandenburger, E., Gressler, M., Leonhardt, R., Lackner, G., Habel, A., Hertweck, C., Brock, M., Hoffmeister, D., 2017. A Highly Conserved Basidiomycete Peptide Synthetase Produces a Trimeric Hydroxamate Siderophore. Appl. Environ. Microbiol. 83, 1–13.

https://doi.org/10.1128/aem.01478-17

Brown, A.S., Calcott, M.J., Owen, J.G., Ackerley, D.F., 2018. Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines. Nat. Prod. Rep. https://doi.org/10.1039/C8NP00036K

Bryksin, A., Matsumura, I., 2010. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. Biotechniques 48, 463–465. https://doi.org/10.2144/000113418

Bumpus, S.B., Evans, B.S., Thomas, P.M., Ntai, I., Kelleher, N.L., 2009. A proteomics approach to discovering natural products and their biosynthetic pathways. Nat. Biotechnol. 27, 951–956. https://doi.org/10.1038/nbt.1565

Bushley, K.E., Ripoll, D.R., Turgeon, B.G., 2008. Module evolution and substrate specificity of fungal nonribosomal peptide synthetases involved in siderophore biosynthesis. BMC Evol. Biol. 8, 328. https://doi.org/10.1186/1471-2148-8-328

Caboche, S., Leclère, V., Pupin, M., Kucherov, G., Jacques, P., 2010. Diversity of monomers in nonribosomal peptides: Towards the prediction of origin and biological activity. J. Bacteriol. 192, 5143–5150. https://doi.org/10.1128/JB.00315-10

Caboche, S., Pupin, M., Leclère, V., Fontaine, A., Jacques, P., Kucherov, G., 2008. NORINE: A database of nonribosomal peptides. Nucleic Acids Res. 36, 326–331. https://doi.org/10.1093/nar/gkm792

Caradec, T., Pupin, M., Vanvlassenbroeck, A., Devignes, M.-D., Smaïl-Tabbone, M., Jacques, P., Leclère, V., 2014. Prediction of Monomer Isomery in Florine: A Workflow Dedicated to Nonribosomal Peptide Discovery. PLoS One 9, e85667. https://doi.org/10.1371/journal.pone.0085667

Carter, M., Shieh, J., 2015. Molecular Cloning and Recombinant DNA Technology, in: Guide to Research Techniques in Neuroscience. Elsevier, pp. 219–237. https://doi.org/10.1016/b978-0-12-800511-8.00010-1

Cavaleiro, A.M., Kim, S.H., Seppälä, S., Nielsen, M.T., Nørholm, M.H.H., 2015. Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning. ACS Synth. Biol. 4, 1042–1046. https://doi.org/10.1021/acssynbio.5b00113

Celińska, E., Ledesma-Amaro, R., Larroude, M., Rossignol, T., Pauthenier, C., Nicaud, J.M., 2017. Golden Gate Assembly system dedicated to complex pathway manipulation in Yarrowia lipolytica. Microb. Biotechnol. 10, 450–455. https://doi.org/10.1111/1751-7915.12605

Challis, G.L., Naismith, J.H., 2004. Structural aspects of non-ribosomal peptide biosynthesis. Curr. Opin. Struct. Biol. https://doi.org/10.1016/j.sbi.2004.10.005

Challis, G.L., Ravel, J., Townsend, C.A., 2000. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem. Biol. 7, 211–24.

Chanda, A., Roze, L. V., Kang, S., Artymovich, K.A., Hicks, G.R., Raikhel, N. V., Calvo, A.M., Linz, J.E., 2009. A key role for vesicles in fungal secondary metabolism. Proc. Natl. Acad. Sci. 106, 19533–19538. https://doi.org/10.1073/pnas.0907416106

Chang, S.L., Chiang, Y.M., Yeh, H.H., Wu, T.K., Wang, C.C.C., 2013. Reconstitution of the early steps of gliotoxin biosynthesis in Aspergillus nidulans reveals the role of the monooxygenase GliC. Bioorganic Med. Chem. Lett. 23, 2155–2157. https://doi.org/10.1016/j.bmcl.2013.01.099

Chavali, A.K., Rhee, S.Y., 2018. Bioinformatics tools for the identification of gene clusters that

biosynthesize specialized metabolites. Brief. Bioinform. 19, 1022–1034. https://doi.org/10.1093/bib/bbx020

Chevrette, M.G., Aicheler, F., Kohlbacher, O., Currie, C.R., Medema, M.H., 2017. SANDPUMA: ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across Actinobacteria. Bioinformatics 33, 3202–3210. https://doi.org/10.1093/bioinformatics/btx400

Chiang, Y.-M., Oakley, C.E., Ahuja, M., Entwistle, R., Schultz, A., Chang, S.-L., Sung, C.T., Wang, C.C.C., Oakley, B.R., 2013. An efficient system for heterologous expression of secondary metabolite genes in Aspergillus nidulans. J. Am. Chem. Soc. 135, 7720–31. https://doi.org/10.1021/ja401945a

Chiang, Y.M., Wang, C.C.C., Oakley, B.R., 2014. Analyzing Fungal Secondary Metabolite Genes and Gene Clusters. Nat. Prod. Discourse, Divers. Des. 171–193. https://doi.org/10.1002/9781118794623.ch10

Chooi, Y.-H., Tang, Y., 2012. Navigating the fungal polyketide chemical space: from genes to molecules. J. Org. Chem. 77, 9933–53. https://doi.org/10.1021/jo301592k

Chutrakul, C., Alcocer, M., Bailey, K., Peberdy, J.F., 2008. The production and characterisation of trichotoxin peptaibols, by Trichoderma asperellum. Chem. Biodivers. 5, 1694–706. https://doi.org/10.1002/cbdv.200890158

Clevenger, K.D., Bok, J.W., Ye, R., Miley, G.P., Verdan, M.H., Velk, T., Chen, C., Yang, K., Robey, M.T., Gao, P., Lamprecht, M., Thomas, P.M., Islam, M.N., Palmer, J.M., Wu, C.C., Keller, N.P., Kelleher, N.L., 2017. A scalable platform to identify fungal secondary metabolites and their gene clusters. Nat. Chem. Biol. 13, 895–901. https://doi.org/10.1038/nchembio.2408

Collins, J., Brüning, H.J., 1978. Plasmids useable as gene-cloning vectors in an in vitro packaging by coliphage λ: "cosmids." Gene 4, 85–107. https://doi.org/10.1016/0378-1119(78)90023-9

Conway, K.R., Boddy, C.N., 2013. ClusterMine360: A database of microbial PKS/NRPS biosynthesis. Nucleic Acids Res. 41, 402–407. https://doi.org/10.1093/nar/gks993

Cundliffe, E., Demain, A.L., 2010. Avoidance of suicide in antibiotic-producing microbes. J. Ind. Microbiol. Biotechnol. https://doi.org/10.1007/s10295-010-0721-x

De Mattos-Shipley, K.M.J., Greco, C., Heard, D.M., Hough, G., Mulholland, N.P., Vincent, J.L., Micklefield, J., Simpson, T.J., Willis, C.L., Cox, R.J., Bailey, A.M., 2018. The cycloaspeptides: Uncovering a new model for methylated nonribosomal peptide biosynthesis. Chem. Sci. 9, 4109–4117. https://doi.org/10.1039/c8sc00717a

Demain, A.L., 2014. Importance of microbial natural products and the need to revitalize their discovery. J. Ind. Microbiol. Biotechnol. 41, 185–201. https://doi.org/10.1007/s10295-013-1325-z

do Valle Matta, M.A., Jonniaux, J.L., Balzi, E., Goffeau, A., van den Hazel, B., 2001. Novel target genes of the yeast regulator Pdr1p: a contribution of the TPO1 gene in resistance to quinidine and other drugs. Gene 272, 111–9.

Dolan, S.K., O'Keeffe, G., Jones, G.W., Doyle, S., 2015. Resistance is not futile: gliotoxin biosynthesis, functionality and utility. Trends Microbiol. 23, 419–28. https://doi.org/10.1016/j.tim.2015.02.005

Du, L., Lou, L., 2010. PKS and NRPS release mechanisms. Nat. Prod. Rep. 27, 255–78. https://doi.org/10.1039/b912037h

1385 Duckworth, B.P., Wilson, D.J., Aldrich, C.C., 2016. Measurement of nonribosomal peptide
1386 synthetase adenylation domain activity using a continuous hydroxylamine release assay, in:
1387 Methods in Molecular Biology. Humana Press Inc., pp. 53–61. https://doi.org/10.1007/978-1-
1388 4939-3375-4_3

1389 Eddy, S.R., 2011. Accelerated Profile HMM Searches. PLoS Comput. Biol. 7, e1002195.
1390 https://doi.org/10.1371/journal.pcbi.1002195

1391 Eddy, S.R., 1998. Profile hidden Markov models. Bioinformatics 14, 755–63.

1392 Elliott, T.A., Gregory, T.R., 2015. What's in a genome? The C-value enigma and the evolution of
1393 eukaryotic genome content. Philos. Trans. R. Soc. B Biol. Sci.
1394 https://doi.org/10.1098/rstb.2014.0331

1395 Engler, C., Kandzia, R., Marillonnet, S., 2008. A one pot, one step, precision cloning method with
1396 high throughput capability. PLoS One 3. https://doi.org/10.1371/journal.pone.0003647

1397 Erler, A., Maresca, M., Fu, J., Stewart, A.F., 2006. Recombineering reagents for improved inducible
1398 expression and selection marker re-use in Schizosaccharomyces pombe. Yeast 23, 813–23.
1399 https://doi.org/10.1002/yea.1396

1400 Esmaeel, Q., Chevalier, M., Chataigné, G., Subashkumar, R., Jacques, P., Leclère, V., 2016.
1401 Nonribosomal peptide synthetase with a unique iterative-alternative-optional mechanism
1402 catalyzes amonabactin synthesis in Aeromonas. Appl. Microbiol. Biotechnol. 100, 8453–8463.
1403 https://doi.org/10.1007/s00253-016-7773-4

1404 Esmaeel, Q., Pupin, M., Jacques, P., Leclère, V., 2017. Nonribosomal peptides and polyketides of
1405 Burkholderia: new compounds potentially implicated in biocontrol and pharmaceuticals.
1406 Environ. Sci. Pollut. Res. 1–14. https://doi.org/10.1007/s11356-017-9166-3

1407 Farag, S., Bleich, R.M., Shank, E.A., Isayev, O., Bowers, A.A., Tropsha, A., 2019. Inter-Modular
1408 Linkers play a crucial role in governing the biosynthesis of non-ribosomal peptides.
1409 Bioinformatics. https://doi.org/10.1093/bioinformatics/btz127

1410 Felenbok, B., Flipphi, M., Nikolaev, I., 2004. Ethanol catabolism in Aspergillus nidulans: A model
1411 system for studying gene regulation 69, 149–204. https://doi.org/10.1016/s0079-
1412 6603(01)69047-0

1413 Fickers, P., Leclère, V., Guez, J.-S., Béchet, M., Coucheney, F., Joris, B., Jacques, P., 2008.
1414 Temperature dependence of mycosubtilin homologue production in Bacillus subtilis
1415 ATCC6633. Res. Microbiol. 159, 449–457. https://doi.org/10.1016/j.resmic.2008.05.004

1416 Fierro, F., García-Estrada, C., Castillo, N.I., Rodríguez, R., Velasco-Conde, T., Martín, J.F., 2006.
1417 Transcriptional and bioinformatic analysis of the 56.8 kb DNA region amplified in tandem
1418 repeats containing the penicillin gene cluster in Penicillium chrysogenum. Fungal Genet. Biol.
1419 43, 618–629. https://doi.org/10.1016/j.fgb.2006.03.001

1420 Fisch, K.M., Bakeer, W., Yakasai, A.A., Song, Z., Pedrick, J., Wasil, Z., Bailey, A.M., Lazarus,
1421 C.M., Simpson, T.J., Cox, R.J., 2011. Rational domain swaps decipher programming in fungal
1422 highly reducing polyketide synthases and resurrect an extinct metabolite. J. Am. Chem. Soc.
1423 133, 16635–16641. https://doi.org/10.1021/ja206914q

1424 Fleming, A., 1929. On the antibacterial action of cultures of a penicillium, with special reference to
1425 their use in the isolation of B. influenzae. Br. J. Exp. Pathol. X, 226–236.
1426 https://doi.org/10.1093/clinids/2.1.129

Freeman, M.F., Vagstad, A.L., Piel, J., 2016. Polytheonamide biosynthesis showcasing the metabolic potential of sponge-associated uncultivated “Entotheonella” bacteria. Curr. Opin. Chem. Biol. 31, 8–14. https://doi.org/10.1016/j.cbpa.2015.11.002

Frisvad, J.C., Rank, C., Nielsen, K.F., Larsen, T.O., 2009. Metabolomics of Aspergillus fumigatus. Med. Mycol. 47, 53–71. https://doi.org/10.1080/13693780802307720

Fujii, I., Ono, Y., Tada, H., Gomi, K., Ebizuka, Y., Sankawa, U., 1996. Cloning of the polyketide synthase gene atX from Aspergillus terreus and its identification as the 6-methylsalicylic acid synthase gene by heterologous expression. Mol. Gen. Genet. 253, 1–10. https://doi.org/10.1007/s004380050289

Gao, X., Haynes, S.W., Ames, B.D., Wang, P., Vien, L.P., Walsh, C.T., Tang, Y., 2012. Cyclization of Fungal Nonribosomal Peptides by a Terminal Condensation-Like Domain. Nat Chem Biol 412, 426–434. https://doi.org/10.1016/j.virol.2011.01.029.The

Gao, X., Jiang, W., Jiménez-Osés, G., Choi, M.S., Houk, K.N., Tang, Y., Walsh, C.T., 2013. An iterative, bimodular nonribosomal peptide synthetase that converts anthranilate and tryptophan into tetracyclic asperlicins. Chem. Biol. 20, 870–878. https://doi.org/10.1016/j.chembiol.2013.04.019

Gardiner, D.M., Howlett, B.J., 2005. Bioinformatic and expression analysis of the putative gliotoxin biosynthetic gene cluster of *Aspergillus fumigatus*. FEMS Microbiol. Lett. 248, 241–248. https://doi.org/10.1016/j.femsle.2005.05.046

Geib, E., Brock, M., 2017. ATNT: an enhanced system for expression of polycistronic secondary metabolite gene clusters in Aspergillus niger. Fungal Biol. Biotechnol. 4, 13. https://doi.org/10.1186/s40694-017-0042-1

Geib, E., Gressler, M., Viediernikova, I., Hillmann, F., Jacobsen, I.D., Nietzsche, S., Hertweck, C., Brock, M., 2016. A Non-canonical Melanin Biosynthesis Pathway Protects Aspergillus terreus Conidia from Environmental Stress. Cell Chem. Biol. 23, 587–597. https://doi.org/10.1016/j.chembiol.2016.03.014

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345. https://doi.org/10.1038/nmeth.1318

Gibson, D.M., Donzelli, B.G.G., Krasnoff, S.B., Keyhani, N.O., 2014. Discovering the secondary metabolite potential encoded within entomopathogenic fungi. Nat. Prod. Rep. 31, 1287–1305. https://doi.org/10.1039/c4np00054d

Gidijala, L., Bovenberg, R. AL, Klaassen, P., van der Klei, I.J., Veenhuis, M., Kiel, J.A.K.W., 2008. Production of functionally active Penicillium chrysogenum isopenicillin N synthase in the yeast Hansenula polymorpha. BMC Biotechnol. 8, 29. https://doi.org/10.1186/1472-6750-8-29

Gidijala, L., Kiel, J.A.K.W., Douma, R.D., Seifar, R.M., van Gulik, W.M., Bovenberg, R.A.L., Veenhuis, M., van der Klei, I.J., 2009. An engineered yeast efficiently secreting penicillin. PLoS One 4. https://doi.org/10.1371/journal.pone.0008317

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., Shabalov, I., 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 42, D699–D704. https://doi.org/10.1093/nar/gkt1183

Guo, C.J., Sun, W.W., Bruno, K.S., Oakley, B.R., Keller, N.P., Wang, C.C.C., 2015. Spatial

regulation of a common precursor from two distinct genes generates metabolite diversity. Chem. Sci. 6, 5913–5921. https://doi.org/10.1039/c5sc01058f

Halo, L.M., Marshall, J.W., Yakasai, A.A., Song, Z., Butts, C.P., Crump, M.P., Heneghan, M., Bailey, A.M., Simpson, T.J., Lazarus, C.M., Cox, R.J., 2008. Authentic heterologous expression of the tenellin iterative polyketide synthase nonribosomal peptide synthetase requires cooexpression with an enoyl reductase. ChemBioChem 9, 585–594. https://doi.org/10.1002/cbic.200700390

Hamada, T., Matsunaga, S., Yano, G., Fusetani, N., 2005. Polytheonamides A and B, highly cytotoxic, linear polypeptides with unprecedented structural features, from the marine sponge, Theonella swinhoei. J. Am. Chem. Soc. 127, 110–8. https://doi.org/10.1021/ja045749e

Harris, D.M., van der Krogt, Z.A., Klaassen, P., Raamsdonk, L.M., Hage, S., van den Berg, M.A., Bovenberg, R.A.L., Pronk, J.T., Daran, J.M., 2009. Exploring and dissecting genome-wide gene expression responses of Penicillium chrysogenum to phenylacetic acid consumption and penicillinG production. BMC Genomics 10. https://doi.org/10.1186/1471-2164-10-75

Harvey, C.J.B., Tang, M., Schlecht, U., Horecka, J., Fischer, C.R., Lin, H.-C., Li, J., Naughton, B., Cherry, J., Miranda, M., Li, Y.F., Chu, A.M., Hennessy, J.R., Vandova, G.A., Inglis, D., Aiyar, R.S., Steinmetz, L.M., Davis, R.W., Medema, M.H., Sattely, E., Khosla, C., St Onge, R.P., Tang, Y., Hillenmeyer, M.E., 2018. HEx: A heterologous expression platform for the discovery of fungal natural products. Sci. Adv. 4, eaar5459. https://doi.org/10.1126/sciadv.aar5459

Haynes, S.W., Gao, X., Tang, Y., Walsh, C.T., 2013. Complexity generation in fungal peptidyl alkaloid biosynthesis: A two-enzyme pathway to the hexacyclic MDR export pump inhibitor ardeemin. ACS Chem. Biol. 8, 741–748. https://doi.org/10.1021/cb3006787

He, Y., Wang, B., Chen, W., Cox, R.J., He, J., Chen, F., 2018. Recent advances in reconstructing microbial secondary metabolites biosynthesis in Aspergillus spp. Biotechnol. Adv. 36, 739–783. https://doi.org/10.1016/j.biotechadv.2018.02.001

Heather, J.M., Chain, B., 2016. The sequence of sequencers: The history of sequencing DNA. Genomics 107, 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

Heneghan, M.N., Yakasai, A.A., Halo, L.M., Song, Z., Bailey, A.M., Simpson, T.J., Cox, R.J., Lazarus, C.M., 2010. First heterologous reconstruction of a complete functional fungal biosynthetic multigene cluster. ChemBioChem 11, 1508–1512. https://doi.org/10.1002/cbic.201000259

Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K., Pease, L.R., 1989. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene 77, 61–8.

Hughes, S.R., Cox, E.J., Bang, S.S., Pinkelman, R.J., López-Núñez, J.C., Saha, B.C., Qureshi, N., Gibbons, W.R., Fry, M.R., Moser, B.R., Bischoff, K.M., Liu, S., Sterner, D.E., Butt, T.R., Riedmuller, S.B., Jones, M.A., Riaño-Herrera, N.M., 2015. Process for Assembly and Transformation into *Saccharomyces cerevisiae* of a Synthetic Yeast Artificial Chromosome Containing a Multigene Cassette to Express Enzymes That Enhance Xylose Utilization Designed for an Automated Platform. J. Lab. Autom. 20, 621–635. https://doi.org/10.1177/2211068215573188

Hühner, E., Backhaus, K., Kraut, R., Li, S.M., 2018. Production of α-keto carboxylic acid dimers in yeast by overexpression of NRPS-like genes from Aspergillus terreus. Appl. Microbiol. Biotechnol. 102, 1663–1672. https://doi.org/10.1007/s00253-017-8719-1

Huo, L., Rachid, S., Stadler, M., Wenzel, S.C., Müller, R., 2012. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. Chem. Biol. 19, 1278–1287. https://doi.org/10.1016/j.chembiol.2012.08.013

Inglis, D.O., Binkley, J., Skrzypek, M.S., Arnaud, M.B., Cerqueira, G.C., Shah, P., Wymore, F., Wortman, J.R., Sherlock, G., 2013. Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of Aspergillus nidulans, A. fumigatus, A. niger and A. oryzae. BMC Microbiol. 13. https://doi.org/10.1186/1471-2180-13-91

Ishiuchi, K., Nakazawa, T., Ookuma, T., Sugimoto, S., Sato, M., Tsunematsu, Y., Ishikawa, N., Noguchi, H., Hotta, K., Moriya, H., Watanabe, K., 2012. Establishing a New Methodology for Genome Mining and Biosynthesis of Polyketides and Peptides through Yeast Molecular Genetics. ChemBioChem 13, 846–854. https://doi.org/10.1002/cbic.201100798

Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D., Dilthey, A.T., Fiddes, I.T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H.E., Pedersen, B.S., Rhie, A., Richardson, H., Quinlan, A.R., Snutch, T.P., Tee, L., Paten, B., Phillippy, A.M., Simpson, J.T., Loman, N.J., Loose, M., 2018. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat. Biotechnol. 36, 338–345. https://doi.org/10.1038/nbt.4060

Jones, E.W., 1991. Tackling the protease problem in Saccharomyces cerevisiae. Methods Enzymol. 194, 428–53.

Kakui, Y., Sunaga, T., Arai, K., Dodgson, J., Ji, L., Csikász-Nagy, A., Carazo-Salas, R., Sato, M., 2015. Module-based construction of plasmids for chromosomal integration of the fission yeast Schizosaccharomyces pombe. Open Biol. 5, 150054. https://doi.org/10.1098/rsob.150054

Kakule, T.B., Lin, Z., Schmidt, E.W., 2014. Combinatorialization of fungal polyketide synthase-peptide synthetase hybrid proteins. J. Am. Chem. Soc. 136, 17882–17890. https://doi.org/10.1021/ja511087p

Kasahara, K., Miyamoto, T., Fujimoto, T., Oguri, H., Tokiwano, T., Oikawa, H., Ebizuka, Y., Fujii, I., 2010. Solanapyrone synthase, a possible Diels-Alderase and iterative type I polyketide synthase encoded in a biosynthetic gene cluster from Alternaria solani. ChemBioChem 11, 1245–1252. https://doi.org/10.1002/cbic.201000173

Kealey, J.T., Liu, L., Santi, D. V, Betlach, M.C., Barr, P.J., 1998. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. Proc. Natl. Acad. Sci. U. S. A. 95, 505–9. https://doi.org/10.1073/pnas.95.2.505

Keller, N.P., 2015. Translating biosynthetic gene clusters into fungal armor and weaponry. Nat. Chem. Biol. https://doi.org/10.1038/nchembio.1897

Khaldi, N., Seifuddin, F.T., Turner, G., Haft, D., Nierman, W.C., Wolfe, K.H., Fedorova, N.D., 2010. SMURF: Genomic mapping of fungal secondary metabolite clusters. Fungal Genet. Biol. https://doi.org/10.1016/j.fgb.2010.06.003

Khayatt, B.I., Overmars, L., Siezen, R.J., Francke, C., 2013. Classification of the adenylation and acyl-transferase activity of NRPS and PKS systems using ensembles of substrate specific hidden Markov models. PLoS One 8, e62136. https://doi.org/10.1371/journal.pone.0062136

Kildgaard, S., Subko, K., Phillips, E., Goidts, V., De La Cruz, M., Díaz, C., Gotfredsen, C.H., Andersen, B., Frisvad, J.C., Nielsen, K.F., Larsen, T.O., 2017. A dereplication and bioguided discovery approach to reveal new compounds from a marine-derived fungus Stilbella fimetaria. Mar. Drugs. https://doi.org/10.3390/md15080253

Kim, J.H., Lee, S.R., Li, L.H., Park, H.J., Park, J.H., Lee, K.Y., Kim, M.K., Shin, B.A., Choi, S.Y., 2011. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. PLoS One 6, e18556. https://doi.org/10.1371/journal.pone.0018556

Knudsen, M., Søndergaard, D., Tofting-Olesen, C., Hansen, F.T., Brodersen, D.E., Pedersen, C.N.S., 2016. Computational discovery of specificity-conferring sites in non-ribosomal peptide synthetases. Bioinformatics 32, 325–9. https://doi.org/10.1093/bioinformatics/btv600

Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P.A., 2019. Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. https://doi.org/10.1038/s41587-019-0072-8

Kopp, F., Marahiel, M.A., 2007. Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. Nat. Prod. Rep. 24, 735. https://doi.org/10.1039/b613652b

Koren, S., Phillippy, A.M., 2015. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. Curr. Opin. Microbiol. 23, 110–120. https://doi.org/10.1016/j.mib.2014.11.014

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. Genome Res. 27, 722–736. https://doi.org/10.1101/gr.215087.116

Kosuri, S., Church, G.M., 2014. Large-scale de novo DNA synthesis: technologies and applications. Nat. Methods 11, 499–507. https://doi.org/10.1038/nmeth.2918

Kouprina, N., Larionov, V., 2016. Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. Chromosoma 125, 621–32. https://doi.org/10.1007/s00412-016-0588-3

Kumakura, N., Ueno, A., Shirasu, K., 2019. Establishment of a selection marker recycling system for sequential transformation of the plant-pathogenic fungus Colletotrichum orbiculare. Mol. Plant Pathol. 20, 447–459. https://doi.org/10.1111/mpp.12766

Kunakom, S., Eustáquio, A.S., 2019. Natural Products and Synthetic Biology: Where We Are and Where We Need To Go. mSystems 4, 1–4. https://doi.org/10.1128/msystems.00113-19

Lawen, A., Zocher, R., 1990. Cyclosporin synthetase. The most complex peptide synthesizing multienzyme polypeptide so far described. J. Biol. Chem. 265, 11355–60.

Lazarus, C.M., Williams, K., Bailey, A.M., 2014. Reconstructing fungal natural product biosynthetic pathways. Nat. Prod. Rep. 31, 1339–1347. https://doi.org/10.1039/c4np00084f

Li, C., Lin, Y., Zheng, X., Yuan, Q., Pang, N., Liao, X., Huang, Y., Zhang, X., Liang, S., 2017. Recycling of a selectable marker with a self-excisable plasmid in Pichia pastoris. Sci. Rep. 7, 11113. https://doi.org/10.1038/s41598-017-11494-5

Li, M.H.T., Ung, P.M.U., Zajkowski, J., Garneau-Tsodikova, S., Sherman, D.H., 2009. Automated genome mining for natural products. BMC Bioinformatics 10, 185. https://doi.org/10.1186/1471-2105-10-185

Li, W., Fan, A., Wang, L., Zhang, P., Liu, Z., An, Z., Yin, W.B., 2018. Asperphenamate biosynthesis reveals a novel two-module NRPS system to synthesize amino acid esters in fungi. Chem. Sci. 9, 2589–2594. https://doi.org/10.1039/c7sc02396k

Lipmann, F., Gevers, W., Kleinkauf, H., Roskoski, R., 1971. Polypeptide synthesis on protein templates: the enzymatic synthesis of gramicidin S and tyrocidine. Adv. Enzymol. Relat. Areas

Mol. Biol. 35, 1–34.

Loman, N.J., Constantinidou, C., Chan, J.Z.M., Halachev, M., Sergeant, M., Penn, C.W., Robinson, E.R., Pallen, M.J., 2012. High-throughput bacterial genome sequencing: An embarrassment of choice, a world of opportunity. Nat. Rev. Microbiol. 10, 599–606. https://doi.org/10.1038/nrmicro2850

Lu, H., Giordano, F., Ning, Z., 2016. Oxford Nanopore MinION Sequencing and Genome Assembly. Genomics. Proteomics Bioinformatics 14, 265–279. https://doi.org/10.1016/j.gpb.2016.05.004

Luo, Y., Cobb, R.E., Zhao, H., 2014. Recent advances in natural product discovery. Curr. Opin. Biotechnol. 30, 230–237. https://doi.org/10.1016/j.copbio.2014.09.002

Luo, Y., Enghiad, B., Zhao, H., 2016. New tools for reconstruction and heterologous expression of natural product biosynthetic gene clusters. Nat. Prod. Rep. 33, 174–82. https://doi.org/10.1039/c5np00085h

Maiya, S., Grundmann, A., Li, S.M., Turner, G., 2006. The fumitremorgin gene cluster of Aspergillus fumigatus: Identification of a gene encoding brevianamide F synthetase. ChemBioChem 7, 1062–1069. https://doi.org/10.1002/cbic.200600003

Manavalan, B., Murugapiran, S.K., Lee, G., Choi, S., 2010. Molecular modeling of the reductase domain to elucidate the reaction mechanism of reduction of peptidyl thioester into its corresponding alcohol in non-ribosomal peptide synthetases. BMC Struct. Biol. 10, 1. https://doi.org/10.1186/1472-6807-10-1

Martín, J.F., Casqueiro, J., Liras, P., 2005. Secretion systems for secondary metabolites: how producer cells send out messages of intercellular communication. Curr. Opin. Microbiol. 8, 282–293. https://doi.org/10.1016/j.mib.2005.04.009

Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., Porro, D., 2012. Recombinant Protein Production in Yeasts, in: Methods in Molecular Biology (Clifton, N.J.). pp. 329–358. https://doi.org/10.1007/978-1-61779-433-9_17

Mattern, D.J., Valiante, V., Unkles, S.E., Brakhage, A.A., 2015. Synthetic biology of fungal natural products. Front. Microbiol. 6, 1–7. https://doi.org/10.3389/fmicb.2015.00775

Maury, J., Germann, S.M., Baallal Jacobsen, S.A., Jensen, N.B., Kildegaard, K.R., Herrgàrd, M.J., Schneider, K., Koza, A., Forster, J., Nielsen, J., Borodina, I., 2016. EasyCloneMulti: A set of vectors for simultaneous and multiple genomic integrations in saccharomyces cerevisiae. PLoS One 11, 1–22. https://doi.org/10.1371/journal.pone.0150394

McErlean, M., Overbay, J., Van Lanen, S., 2019. Refining and expanding nonribosomal peptide synthetase function and mechanism. J. Ind. Microbiol. Biotechnol. 46, 493–513. https://doi.org/10.1007/s10295-018-02130-w

McGinnis, S., Madden, T.L., 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 32, W20-5. https://doi.org/10.1093/nar/gkh435

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., Breitling, R., 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 39, W339-46. https://doi.org/10.1093/nar/gkr466

Medema, M.H., Takano, E., Breitling, R., 2013. Detecting Sequence Homology at the Gene Cluster Level with MultiGeneBlast. Mol. Biol. Evol. 30, 1218–1223. https://doi.org/10.1093/molbev/mst025

Meyer, V., Wanka, F., van Gent, J., Arentshorst, M., van den Hondel, C.A.M.J.J., Ram, A.F.J., 2011. Fungal Gene Expression on Demand: an Inducible, Tunable, and Metabolism-Independent Expression System for Aspergillus niger. Appl. Environ. Microbiol. 77, 2975–2983. https://doi.org/10.1128/aem.02740-10

Minowa, Y., Araki, M., Kanehisa, M., 2007. Comprehensive Analysis of Distinctive Polyketide and Nonribosomal Peptide Structural Motifs Encoded in Microbial Genomes. J. Mol. Biol. 368, 1500–1517. https://doi.org/10.1016/J.JMB.2007.02.099

Mitchell, L.A., Chuang, J., Agmon, N., Khunsriraksakul, C., Phillips, N.A., Cai, Y., Truong, D.M., Veerakumar, A., Wang, Y., Mayorga, M., Blomquist, P., Sadda, P., Trueheart, J., Boeke, J.D., 2015. Versatile genetic assembly system (VEGAS) to assemble pathways for expression in S. cerevisiae. Nucleic Acids Res. 43, 6620–30. https://doi.org/10.1093/nar/gkv466

Mofid, M.R., Finking, R., Marahiel, M.A., 2002. Recognition of hybrid peptidyl carrier proteins/acyl carrier proteins in nonribosomal peptide synthetase modules by the 4′-phophopantetheinyl transferases AcpS and Sfp. J. Biol. Chem. 277, 17023–17031. https://doi.org/10.1074/jbc.M200120200

Mootz, H.D., Schörgendorfer, K., Marahiel, M.A., 2002. Functional characterization of 4′-phosphopantetheinyl transferase genes of bacterial and fungal origin by complementation of Saccharomyces cerevisiae lys5. FEMS Microbiol. Lett. 213, 51–57. https://doi.org/10.1016/S0378-1097(02)00760-7

Müller, W.H., van der Krift, T.P., Krouwer, A.J., Wösten, H.A., van der Voort, L.H., Smaal, E.B., Verkleij, A.J., 2018. Localization of the pathway of the penicillin biosynthesis in Penicillium chrysogenum. EMBO J. 10, 489–495. https://doi.org/10.1002/j.1460-2075.1991.tb07971.x

Munawar, A., Marshall, J.W., Cox, R.J., Bailey, A.M., Lazarus, C.M., 2013. Isolation and Characterisation of a Ferrirhodin Synthetase Gene from the Sugarcane Pathogen Fusarium sacchari. ChemBioChem 14, 388–394. https://doi.org/10.1002/cbic.201200587

Mushtaq, S., Abbasi, B.H., Uzair, B., Abbasi, R., 2018. Natural products as reservoirs of novel therapeutic agents. EXCLI J. 17, 420–451. https://doi.org/10.17179/excli2018-1174

Nah, H.-J., Pyeon, H.-R., Kang, S.-H., Choi, S.-S., Kim, E.-S., 2017. Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in Streptomyces Species. Front. Microbiol. 8. https://doi.org/10.3389/fmicb.2017.00394

NCBI Resource Coordinators, 2016. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 44, D7-19. https://doi.org/10.1093/nar/gkv1290

Newbert, R W, Barton, B., Greaves, P., Harper, J., Turner, G., 1997. Analysis of a commercially improved Penicillium chrysogenum strain series: involvement of recombinogenic regions in amplification and deletion of the penicillin biosynthesis gene cluster. J. Ind. Microbiol. Biotechnol. 19, 18–27.

Newbert, R. W., Barton, B., Greaves, P., Harper, J., Turner, G., 1997. Analysis of a commercially improved Penicillium chrysogenum strain series: Involvement of recombinogenic regions in amplification and deletion of the penicillin biosynthesis gene cluster. J. Ind. Microbiol. Biotechnol. 19, 18–27. https://doi.org/10.1038/sj.jim.2900411

Nielsen, J.C., Nielsen, J., 2017. Development of fungal cell factories for the production of secondary metabolites: Linking genomics and metabolism. Synth. Syst. Biotechnol. 2, 5–12. https://doi.org/10.1016/j.synbio.2017.02.002

1684 Nielsen, M.L., Isbrandt, T., Petersen, L.M., Mortensen, U.H., Andersen, M.R., Hoof, J.B., Larsen,
1685 T.O., 2016. Linker flexibility facilitates module exchange in fungal hybrid PKS-NRPS
1686 engineering. PLoS One 11, 1–18. https://doi.org/10.1371/journal.pone.0161199

1687 Ongley, S.E., Bian, X., Neilan, B.A., Müller, R., 2013. Recent advances in the heterologous
1688 expression of microbial natural product biosynthetic pathways. Nat. Prod. Rep. 30, 1121–1138.
1689 https://doi.org/10.1039/c3np70034h

1690 Pahirulzaman, K.A.K., Williams, K., Lazarus, C.M., 2012. A toolkit for heterologous expression of
1691 metabolic pathways in aspergillus oryzae, 1st ed, Methods in Enzymology. Elsevier Inc.
1692 https://doi.org/10.1016/B978-0-12-404634-4.00012-7

1693 Pan, R., Zhang, J., Shen, W.L., Tao, Z.Q., Li, S.P., Yan, X., 2011. Sequential deletion of Pichia
1694 pastoris genes by a self-excisable cassette. FEMS Yeast Res. 11, 292–298.
1695 https://doi.org/10.1111/j.1567-1364.2011.00716.x

1696 Phelan, V. V., Du, Y., McLean, J.A., Bachmann, B.O., 2009. Adenylation Enzyme Characterization
1697 Using γ -18O4-ATP Pyrophosphate Exchange. Chem. Biol. 16, 473–478.
1698 https://doi.org/10.1016/j.chembiol.2009.04.007

1699 Pickens, L., Tang, Y., Chooi, Y.-H., 2014. Metabolic Engineering for the Production of Natural
1700 Products. Annu. Rev. Chem. Biomol. Eng. 2, 211–236. https://doi.org/10.1146/annurev-
1701 chembioeng-061010-114209.Metabolic

1702 Prieto, C., García-Estrada, C., Lorenzana, D., Martín, J.F., 2012. NRPSsp: non-ribosomal peptide
1703 synthase substrate predictor. Bioinformatics 28, 426–427.
1704 https://doi.org/10.1093/bioinformatics/btr659

1705 Pupin, M., Esmaeel, Q., Flissi, A., Dufresne, Y., Jacques, P., Leclère, V., 2016. Norine: A powerful
1706 resource for novel nonribosomal peptide discovery. Synth. Syst. Biotechnol. 1, 89–94.
1707 https://doi.org/10.1016/j.synbio.2015.11.001

1708 Quadri, L.E.N., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P., Walsh, C.T., 1998.
1709 Characterization of Sfp , a Bacillus subtilis Phosphopantetheinyl Transferase for Peptidyl
1710 Carrier Protein Domains in Peptide Synthetases Characterization of Sfp , a Bacillus subtilis
1711 Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in. Society 37, 1585–
1712 1595. https://doi.org/10.1021/bi9719861

1713 Qun Yue, L.C.Y.L.G.F.B.X.Z.M.X.S.L.Y.C.C.W.X.N.Z.A.X.L., 2015. Functional Operons in
1714 Secondary Metabolic Gene Clusters in Glarea lozoyensis (Fungi, Ascomycota, Leotiomycetes).
1715 MBio 6.

1716 Rausch, C., Hoof, I., Weber, T., Wohlleben, W., Huson, D.H., 2007. Phylogenetic analysis of
1717 condensation domains in NRPS sheds light on their functional evolution. BMC Evol. Biol. 7, 1–
1718 15. https://doi.org/10.1186/1471-2148-7-78

1719 Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., Huson, D.H., 2005. Specificity prediction of
1720 adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support
1721 vector machines (TSVMs). Nucleic Acids Res. 33, 5799–5808.
1722 https://doi.org/10.1093/nar/gki885

1723 Rhoads, A., Au, K.F., 2015. PacBio Sequencing and Its Applications. Genomics. Proteomics
1724 Bioinformatics 13, 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

1725 Richter, L., Wanka, F., Boecker, S., Storm, D., Kurt, T., Vural, Ö., Süßmuth, R., Meyer, V., 2014.
1726 Engineering of Aspergillus niger for the production of secondary metabolites. Fungal Biol.

Biotechnol. 1, 4. https://doi.org/10.1186/s40694-014-0004-9

Rodriguez, S., Kirby, J., Denby, C.M., Keasling, J.D., 2014. Production and quantification of sesquiterpenes in Saccharomyces cerevisiae, including extraction, detection and quantification of terpene products and key related metabolites. Nat. Protoc. 9, 1980–96. https://doi.org/10.1038/nprot.2014.132

Röttig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C., Kohlbacher, O., 2011. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res. 39, W362–W367. https://doi.org/10.1093/nar/gkr323

Sakai, K., Kinoshita, H., Shimizu, T., Nihira, T., 2008. Construction of a Citrinin Gene Cluster Expression System in Heterologous Aspergillus oryzae. J. Biosci. Bioeng. 106, 466–472. https://doi.org/10.1263/jbb.106.466

Salo, O., Guzmán-Chávez, F., Ries, M.I., Lankhorst, P.P., Bovenberg, R.A.L., Vreeken, R.J., Driessen, A.J.M., 2016. Identification of a polyketide synthase involved in sorbicillin biosynthesis by Penicillium chrysogenum. Appl. Environ. Microbiol. 82, 3971–3978. https://doi.org/10.1128/AEM.00350-16

Salo, O. V., Ries, M., Medema, M.H., Lankhorst, P.P., Vreeken, R.J., Bovenberg, R.A.L., Driessen, A.J.M., 2015. Genomic mutational analysis of the impact of the classical strain improvement program on β-lactam producing Penicillium chrysogenum. BMC Genomics 16. https://doi.org/10.1186/s12864-015-2154-4

Samel, S.A., Schoenafinger, G., Knappe, T.A., Marahiel, M.A., Essen, L.-O., 2007. Structural and Functional Insights into a Peptide Bond-Forming Bidomain from a Nonribosomal Peptide Synthetase. Structure 15, 781–792. https://doi.org/10.1016/j.str.2007.05.008

Saraya, R., Gidijala, L., Veenhuis, M., van der Klei, I.J., 2014. Tools for genetic engineering of the yeast Hansenula polymorpha. Methods Mol. Biol. 1152, 43–62. https://doi.org/10.1007/978-1-4939-0563-8_3

Sato, M., Yagishita, F., Mino, T., Uchiyama, N., Patel, A., Chooi, Y.H., Goda, Y., Xu, W., Noguchi, H., Yamamoto, T., Hotta, K., Houk, K.N., Tang, Y., Watanabe, K., 2015. Involvement of Lipocalin-like CghA in Decalin-Forming Stereoselective Intramolecular [4+2] Cycloaddition. ChemBioChem 16, 2294–2298. https://doi.org/10.1002/cbic.201500386

Sauer, B., 1987. Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae. Mol. Cell. Biol. 7, 2087–96.

Schafhauser, T., Kirchner, N., Kulik, A., Huijbers, M.M.E., Flor, L., Caradec, T., Fewer, D.P., Gross, H., Jacques, P., Jahn, L., Jokela, J., Leclère, V., Ludwig-Müller, J., Sivonen, K., van Berkel, W.J.H., Weber, T., Wohlleben, W., van Pée, K.-H., 2016. The cyclochlorotine mycotoxin is produced by the nonribosomal peptide synthetase CctN in Talaromyces islandicus (‘ Penicillium islandicum ’). Environ. Microbiol. 18, 3728–3741. https://doi.org/10.1111/1462-2920.13294

Schauwecker, U.K. and F., 2003. Combinatorial Biosynthesis of Non-Ribosomal Peptides. Comb. Chem. High Throughput Screen. https://doi.org/http://dx.doi.org/10.2174/138620703106298707

Schimming, O., Fleischhacker, F., Nollmann, F.I., Bode, H.B., 2014. Yeast homologous recombination cloning leading to the novel peptides ambactin and xenolindicin. Chembiochem 15, 1290–4. https://doi.org/10.1002/cbic.201402065

Schümann, J., Hertweck, C., 2006. Advances in cloning, functional analysis and heterologous

1769 expression of fungal polyketide synthase genes. J. Biotechnol. 124, 690–703.
1770 https://doi.org/10.1016/j.jbiotec.2006.03.046

1771 Schwarzer, D., Mootz, H.D., Marahiel, M.A., 2001. Exploring the impact of different thioesterase
1772 domains for the design of hybrid peptide synthetases. Chem. Biol. 8, 997–1010.
1773 https://doi.org/10.1016/S1074-5521(01)00068-0

1774 Schwecke, T., Göttling, K., Durek, P., Dueñas, I., Käufer, N.F., Zock-Emmenthal, S., Staub, E.,
1775 Neuhof, T., Dieckmann, R., Von Döhren, H., 2006. Nonribosomal peptide synthesis in
1776 Schizosoccharomyces pombe and the architectures of ferrichrome-type siderophore synthetases
1777 in fungi. ChemBioChem 7, 612–622. https://doi.org/10.1002/cbic.200500301

1778 Setubal, J.C., Almeida, N.F., Wattam, A.R., 2018. Comparative genomics for prokaryotes, in:
1779 Methods in Molecular Biology. https://doi.org/10.1007/978-1-4939-7463-4_3

1780 Sharma, C., Chowdhary, A., 2017. Molecular bases of antifungal resistance in filamentous fungi. Int.
1781 J. Antimicrob. Agents 50, 607–616. https://doi.org/10.1016/j.ijantimicag.2017.06.018

1782 Siewers, V., Chen, X., Huang, L., Zhang, J., Nielsen, J., 2009. Heterologous production of non-
1783 ribosomal peptide LLD-ACV in Saccharomyces cerevisiae. Metab. Eng. 11, 391–397.
1784 https://doi.org/10.1016/j.ymben.2009.08.002

1785 Siewers, V., San-Bento, R., Nielsen, J., 2010. Implementation of communication-mediating domains
1786 for non-ribosomal peptide production in Saccharomyces cerevisiae. Biotechnol. Bioeng. 106,
1787 841–844. https://doi.org/10.1002/bit.22739

1788 Singh, M.P., Leighton, M.M., Barbieri, L.R., Roll, D.M., Urbance, S.E., Hoshan, L., McDonald,
1789 L.A., 2010. Fermentative production of self-toxic fungal secondary metabolites. J. Ind.
1790 Microbiol. Biotechnol. 37, 335–40. https://doi.org/10.1007/s10295-009-0678-9

1791 Sivanathan, S., Scherkenbeck, J., 2014. Cyclodepsipeptides: A Rich Source of Biologically Active
1792 Compounds for Drug Research. Molecules 19, 12368–12420.
1793 https://doi.org/10.3390/molecules190812368

1794 Skiba, M.A., Maloney, F.P., Dan, Q., Fraley, A.E., Aldrich, C.C., Smith, J.L., Brown, W.C., 2018.
1795 PKS–NRPS Enzymology and Structural Biology: Considerations in Protein Production, 1st ed,
1796 Methods in Enzymology. Elsevier Inc. https://doi.org/10.1016/bs.mie.2018.01.035

1797 Smith, C., Day, P.J., Walker, M.R., 1993. Generation of cohesive ends on PCR products by UDG-
1798 mediated excision of dU, and application for cloning into restriction digest-linearized vectors.
1799 PCR Methods Appl. 2, 328–32.

1800 Smith, D.J., Burnham, M.K., Bull, J.H., Hodgson, J.E., Ward, J.M., Browne, P., Brown, J., Barton,
1801 B., Earl, A.J., Turner, G., 1990. Beta-lactam antibiotic biosynthetic genes have been conserved
1802 in clusters in prokaryotes and eukaryotes. EMBO J. 9, 741–7.

1803 Song, Z., Bakeer, W., Marshall, J.W., Yakasai, A.A., Khalid, R.M., Collemare, J., Skellam, E.,
1804 Tharreau, D., Lebrun, M.H., Lazarus, C.M., Bailey, A.M., Simpson, T.J., Cox, R.J., 2015.
1805 Heterologous expression of the avirulence gene ACE1 from the fungal rice pathogen
1806 Magnaporthe oryzae. Chem. Sci. 6, 4837–4845. https://doi.org/10.1039/c4sc03707c

1807 Soukup, A.A., Keller, N.P., Wiemann, P., 2016. Nonribosomal Peptide and Polyketide Biosynthesis
1808 1401, 149–160. https://doi.org/10.1007/978-1-4939-3375-4

1809 Spröte, P., Brakhage, A.A., Hynes, M.J., 2009. Contribution of Peroxisomes to Penicillin
1810 Biosynthesis in Aspergillus nidulans . Eukaryot. Cell 8, 421–423.
1811 https://doi.org/10.1128/ec.00374-08

1812 Stachelhaus, T., Mootz, H.D., Marahiel, M.A., 1999. The specificity-conferring code of adenylation
1813 domains in nonribosomal peptide synthetases. Chem. Biol. 6, 493–505.
1814 https://doi.org/10.1016/S1074-5521(99)80082-9

1815 Stajich, J.E., 2017. Fungal Genomes and Insights into the Evolution of the Kingdom, in: The Fungal
1816 Kingdom. American Society of Microbiology, pp. 619–633.
1817 https://doi.org/10.1128/microbiolspec.FUNK-0055-2016

1818 Starcevic, A., Zucko, J., Simunkovic, J., Long, P.F., Cullum, J., Hranueli, D., 2008. ClustScan: an
1819 integrated program package for the semi-automatic annotation of modular biosynthetic gene
1820 clusters and in silico prediction of novel chemical structures. Nucleic Acids Res. 36, 6882–92.
1821 https://doi.org/10.1093/nar/gkn685

1822 Steiniger, C., Hoffmann, S., Mainz, A., Kaiser, M., Voigt, K., Meyer, V., Süssmuth, R.D., 2017.
1823 Harnessing fungal nonribosomal cyclodepsipeptide synthetases for mechanistic insights and
1824 tailored engineering. Chem. Sci. 8, 7834–7843. https://doi.org/10.1039/c7sc03093b

1825 Stöckmann, C., Scheidle, M., Dittrich, B., Merckelbach, A., Hehmann, G., Melmer, G., Klee, D.,
1826 Büchs, J., Kang, H.A., Gellissen, G., 2009. Process development in Hansenula polymorpha and
1827 Arxula adeninivorans, a re-assessment. Microb. Cell Fact. 8, 22. https://doi.org/10.1186/1475-
1828 2859-8-22

1829 Suárez, T., Peñalva, M.A., 1996. Characterization of a Penicillium chrysogenum gene encoding a
1830 PacC transcription factor and its binding sites in the divergent pcbAB-pcbC promoter of the
1831 penicillin biosynthetic cluster. Mol. Microbiol. 20, 529–540. https://doi.org/10.1046/j.1365-
1832 2958.1996.5421065.x

1833 Sung, C.T., Chang, S.-L., Entwistle, R., Ahn, G., Lin, T.-S., Petrova, V., Yeh, H.-H., Praseuth, M.B.,
1834 Chiang, Y.-M., Oakley, B.R., Wang, C.C.C., 2017. Overexpression of a three-gene conidial
1835 pigment biosynthetic pathway in Aspergillus nidulans reveals the first NRPS known to acetylate
1836 tryptophan. Fungal Genet. Biol. 101, 1–6. https://doi.org/10.1016/j.fgb.2017.01.006

1837 Süssmuth, R.D., Mainz, A., 2017. Nonribosomal Peptide Synthesis—Principles and Prospects.
1838 Angew. Chemie - Int. Ed. 56, 3770–3821. https://doi.org/10.1002/anie.201609079

1839 Tanaka, M., Tokuoka, M., Gomi, K., 2014. Effects of codon optimization on the mRNA levels of
1840 heterologous genes in filamentous fungi. Appl. Microbiol. Biotechnol. 98, 3859–3867.
1841 https://doi.org/10.1007/s00253-014-5609-7

1842 Tang, M.C., Zou, Y., Yee, D., Tang, Y., 2018. Identification of the pyranonigrin A biosynthetic gene
1843 cluster by genome mining in Penicillium thymicola IBT 5891. AIChE J. 64, 4182–4186.
1844 https://doi.org/10.1002/aic.16324

1845 Tarry, M.J., Haque, A.S., Bui, K.H., Schmeing, T.M., 2017. X-Ray Crystallography and Electron
1846 Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a
1847 Flexible Architecture. Structure 25, 783-793.e4. https://doi.org/10.1016/j.str.2017.03.014

1848 Tatusova, T., Ciufo, S., Fedorov, B., O'Neill, K., Tolstoy, I., 2014. RefSeq microbial genomes
1849 database: new representation and annotation strategy. Nucleic Acids Res. 42, D553–D559.
1850 https://doi.org/10.1093/nar/gkt1274

1851 Terfrüchte, M., Joehnk, B., Fajardo-Somera, R., Braus, G.H., Riquelme, M., Schipper, K.,
1852 Feldbrügge, M., 2014. Establishing a versatile Golden Gate cloning system for genetic
1853 engineering in fungi. Fungal Genet. Biol. 62, 1–10. https://doi.org/10.1016/j.fgb.2013.10.012

1854 Tobias, N.J., Ahrendt, T., Schell, U., Miltenberger, M., Hilbi, H., Bode, H.B., 2016. *Legionella*
1855 shows a diverse secondary metabolism dependent on a broad spectrum Sfp-type
1856 phosphopantetheinyl transferase. PeerJ 4, e2720. https://doi.org/10.7717/peerj.2720

1857 Tran, P.N., Yen, M.R., Chiang, C.Y., Lin, H.C., Chen, P.Y., 2019. Detecting and prioritizing
1858 biosynthetic gene clusters for bioactive compounds in bacteria and fungi. Appl. Microbiol.
1859 Biotechnol. https://doi.org/10.1007/s00253-019-09708-z

1860 Tsunematsu, Y., Ishikawa, N., Wakana, D., Goda, Y., Noguchi, H., Moriya, H., Hotta, K., Watanabe,
1861 K., 2013a. Distinct mechanisms for spiro-carbon formation reveal biosynthetic pathway
1862 crosstalk. Nat. Chem. Biol. 9, 818–825. https://doi.org/10.1038/nchembio.1366

1863 Tsunematsu, Y., Ishiuchi, K., Hotta, K., Watanabe, K., 2013b. Yeast-based genome mining,
1864 production and mechanistic studies of the biosynthesis of fungal polyketide and peptide natural
1865 products. Nat. Prod. Rep. 30, 1139–49. https://doi.org/10.1039/c3np70037b

1866 Unkles, S.E., Valiante, V., Mattern, D.J., Brakhage, A.A., 2014. Synthetic biology tools for
1867 bioprospecting of natural products in eukaryotes. Chem. Biol. 21, 502–508.
1868 https://doi.org/10.1016/j.chembiol.2014.02.010

1869 van Dijk, J.W.A., Wang, C.C.C., 2016. Heterologous Expression of Fungal Secondary Metabolite
1870 Pathways in the Aspergillus nidulans Host System, in: Methods in Enzymology.
1871 https://doi.org/10.1016/bs.mie.2016.02.021

1872 Vassaux, A., Tarayre, C., Arguëlles-Arias, A., Compère, P., Delvigne, F., Fickers, P., Jahn, L., Lang,
1873 A., Leclère, V., Ludwig-Müller, J., Ongena, M., Schafhauser, T., Telek, S., Théatre, A., Berkel,
1874 W.J.H., Vandenbol, M., Pée, K., Willems, L., Wohlleben, W., Jacques, P., 2019. Astin C
1875 production by the endophytic fungus *Cyanodermella asteris* in planktonic and immobilized
1876 culture conditions. Biotechnol. J. 1800624. https://doi.org/10.1002/biot.201800624

1877 Vesth, T.C., Brandl, J., Andersen, M.R., 2016. FunGeneClusterS: Predicting fungal gene clusters
1878 from genome and transcriptome data. Synth. Syst. Biotechnol.
1879 https://doi.org/10.1016/j.synbio.2016.01.002

1880 Vournakis, J.N., Elander, R.P., 1983. Genetic manipulation of antibiotic-producing microorganisms.
1881 Science 219, 703–9.

1882 Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q.,
1883 Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: An Integrated Tool for Comprehensive
1884 Microbial Variant Detection and Genome Assembly Improvement. PLoS One 9, e112963.
1885 https://doi.org/10.1371/journal.pone.0112963

1886 Wang, M., Zhao, H., 2014. Characterization and engineering of the adenylation domain of a NRPS-
1887 like protein: A potential biocatalyst for aldehyde generation. ACS Catal. 4, 1219–1225.
1888 https://doi.org/10.1021/cs500039v

1889 Wang, X., Zhang, Xiaoling, Liu, L., Xiang, M., Wang, W., Sun, X., Che, Y., Guo, Liangdong, Liu,
1890 G., Guo, Liyun, Wang, C., Yin, W.B., Stadler, M., Zhang, Xinyu, Liu, X., 2015. Genomic and
1891 transcriptomic analysis of the endophytic fungus reveals its lifestyle and high potential for
1892 synthesis of natural products. BMC Genomics 16, 1–13. https://doi.org/10.1186/s12864-014-
1893 1190-9

1894 Wasil, Z., Pahirulzaman, K.A.K., Butts, C., Simpson, T.J., Lazarus, C.M., Cox, R.J., 2013. One
1895 pathway, many compounds: Heterologous expression of a fungal biosynthetic pathway reveals
1896 its intrinsic potential for diversity. Chem. Sci. 4, 3845–3856. https://doi.org/10.1039/c3sc51785c

Weber, T., 2014. In silico tools for the analysis of antibiotic biosynthetic pathways. Int. J. Med. Microbiol. 304, 230–235. https://doi.org/10.1016/j.ijmm.2014.02.001

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Müller, R., Wohlleben, W., Breitling, R., Takano, E., Medema, M.H., 2015a. antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 43, W237-43. https://doi.org/10.1093/nar/gkv437

Weber, T., Charusanti, P., Musiol-Kroll, E.M., Jiang, X., Tong, Y., Kim, H.U., Lee, S.Y., 2015b. Metabolic engineering of antibiotic factories: New tools for antibiotic production in actinomycetes. Trends Biotechnol. https://doi.org/10.1016/j.tibtech.2014.10.009

Weber, T., Kim, H.U., 2016. The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. Synth. Syst. Biotechnol. 1, 69–79. https://doi.org/10.1016/j.synbio.2015.12.002

Wehrs, M., Prahl, J.-P., Moon, J., Li, Y., Tanjore, D., Keasling, J.D., Pray, T., Mukhopadhyay, A., 2018. Production efficiency of the bacterial non-ribosomal peptide indigoidine relies on the respiratory metabolic state in S. cerevisiae. Microb. Cell Fact. 17, 193. https://doi.org/10.1186/s12934-018-1045-1

Wilson, M.C., Mori, T., Rückert, C., Uria, A.R., Helf, M.J., Takada, K., Gernert, C., Steffens, U.A.E., Heycke, N., Schmitt, S., Rinke, C., Helfrich, E.J.N., Brachmann, A.O., Gurgui, C., Wakimoto, T., Kracht, M., Crüsemann, M., Hentschel, U., Abe, I., Matsunaga, S., Kalinowski, J., Takeyama, H., Piel, J., 2014. An environmental bacterial taxon with a large and distinct metabolic repertoire. Nature 506, 58–62. https://doi.org/10.1038/nature12959

Winn, M., Fyans, J.K., Zhuo, Y., Micklefield, J., 2016. Recent advances in engineering nonribosomal peptide assembly lines. Nat. Prod. Rep. 33, 317–347. https://doi.org/10.1039/c5np00099h

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

Wunsch, C., Mundt, K., Li, S.M., 2015. Targeted production of secondary metabolites by coexpression of non-ribosomal peptide synthetase and prenyltransferase genes in Aspergillus. Appl. Microbiol. Biotechnol. https://doi.org/10.1007/s00253-015-6490-8

Xu, W., Cai, X., Jung, M.E., Tang, Y., 2010. Analysis of intact and dissected fungal polyketide synthase-nonribosomal peptide synthetase in vitro and in saccharomyces cerevisiae. J. Am. Chem. Soc. 132, 13604–13607. https://doi.org/10.1021/ja107084d

Xue, Y., Kong, C., Shen, W., Bai, C., Ren, Y., Zhou, X., Zhang, Y., Cai, M., 2017. Methylotrophic yeast Pichia pastoris as a chassis organism for polyketide synthesis via the full citrinin biosynthetic pathway. J. Biotechnol. 242, 64–72. https://doi.org/10.1016/j.jbiotec.2016.11.031

Yaegashi, J., Oakley, B.R., Wang, C.C.C., 2014. Recent advances in genome mining of secondary metabolite biosynthetic gene clusters and the development of heterologous expression systems in Aspergillus nidulans. J Ind Microbiol Biotechnol 41, 433–442. https://doi.org/10.1007/s10295-013-1386-z.Recent

Yandell, M., Ence, D., 2012. A beginner's guide to eukaryotic genome annotation. Nat. Rev. Genet. 13, 329–342. https://doi.org/10.1038/nrg3174

Yeh, H.-H., Chiang, Y.-M., Entwistle, R., Ahuja, M., Lee, K.-H., Bruno, K.S., Wu, T.-K., Oakley,

1939 B.R., Wang, C.C.C., 2012. Molecular genetic analysis reveals that a nonribosomal peptide
1940 synthetase-like (NRPS-like) gene in Aspergillus nidulans is responsible for microperfuranone
1941 biosynthesis. Appl. Microbiol. Biotechnol. 96, 739–48. https://doi.org/10.1007/s00253-012-
1942 4098-9

1943 Yoshimi, A., Yamaguchi, S., Fujioka, T., Kawai, K., Gomi, K., Machida, M., Abe, K., 2018.
1944 Heterologous production of a novel cyclic peptide compound, KK-1, in Aspergillus oryzae.
1945 Front. Microbiol. 9, 1–12. https://doi.org/10.3389/fmicb.2018.00690

1946 Yu, D., Xu, F., Zhang, S., Zhan, J., 2017. Decoding and reprogramming fungal iterative
1947 nonribosomal peptide synthetases. Nat. Commun. 8, 1–11.
1948 https://doi.org/10.1038/ncomms15349

1949 Yu, D., Xu, F., Zi, J., Wang, S., Gage, D., Zeng, J., Zhan, J., 2013. Engineered production of fungal
1950 anticancer cyclooligomer depsipeptides in Saccharomyces cerevisiae. Metab. Eng. 18, 60–68.
1951 https://doi.org/10.1016/j.ymben.2013.04.001

1952 Zhang, H., Boghigian, B.A., Armando, J., Pfeifer, B.A., 2011. Methods and options for the
1953 heterologous production of complex natural products. Nat. Prod. Rep. 28, 125–151.
1954 https://doi.org/10.1039/c0np00037j

1955 Zhang, J.J., Yamanaka, K., Tang, X., Moore, B.S., 2019. Direct cloning and heterologous expression
1956 of natural product biosynthetic gene clusters by transformation-associated recombination, in:
1957 Methods in Enzymology. Academic Press Inc., pp. 87–110.
1958 https://doi.org/10.1016/bs.mie.2019.02.026

1959 Zhang, W., Liu, J., 2016. Recent Advances in Understanding and Engineering Polyketide Synthesis.
1960 F1000Research 5, 208. https://doi.org/10.12688/f1000research.7326.1

1961 Ziemert, N., Alanjary, M., Weber, T., 2016. The evolution of genome mining in microbes - a review.
1962 Nat. Prod. Rep. 33, 988–1005. https://doi.org/10.1039/c6np00025h

1963 Ziemert, N., Podell, S., Penn, K., Badger, J.H., Allen, E., Jensen, P.R., 2012. The natural product
1964 domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite
1965 gene diversity. PLoS One 7, e34064. https://doi.org/10.1371/journal.pone.0034064

1966 Zobel, S., Boecker, S., Kulke, D., Heimbach, D., Meyer, V., Süssmuth, R.D., 2016. Reprogramming
1967 the Biosynthesis of Cyclodepsipeptide Synthetases to Obtain New Enniatins and Beauvericins.
1968 ChemBioChem 17, 283–287. https://doi.org/10.1002/cbic.201500649

1969

1970 **Table 1:** List of condensation domain subtypes and their associated reactions

Condensation domain subtype	Reaction description	Reference
$^{L}C_{L}$	Bind $_{L}$-amino acids to $_{L}$-amino acids	
$^{D}C_{L}$	Bind $_{D}$-amino acids to $_{L}$-amino acids	
Starter C	Load a fatty acid as first substrate of the synthesis	(Rausch et al., 2007; Ziemert et al., 2012)
Dual C/E	Combine condensation and epimerization activities	
Epimerization	Invert an $_{L}$-form amino acid to its $_{D}$-form enantiomer	
Heterocyclization	Catalyse peptide bond formation, followed by cyclization	
C_{term}	Replace thioesterase activity for the final step of peptide release	(Gao et al., 2012)
Hybrid C	Bind an amino acid to an aminated polyketide in hybrid PKS-NRPS systems	(Ziemert et al., 2012)

1971
1972

1973 **Table 2:** Previous studies reporting NRP heterologous production in yeasts

Natural product (NRP)	NRPS	NRPS MW (kDa)	Native producer	Heterologous host	Expression system	Reference
ACV	ACVS (or PcbAB)	450	*P. chrysogenum*	*H. polymorpha*	Plasmid pZ4 Plasmid pG4U	Gidijala, 2008
				S. cerevisiae	Plasmid pESC Genomic integration	Siewers, 2009
					Plasmid pESC	Awan, 2017
Fumiquinazoline F	Afu6g12080	450	*A. fumigatus*	*S. cerevisiae*	Plasmid pkW1810	Ishiuchi, 2012
	TqaA		*P. aethiopicum*		Plasmid pXW55	Gao, 2012
Tricyclic ardeemin F-Q	ArdA	432	*N. fischeri*	*S. cerevisiae*	Plasmid pXW55	Haynes, 2013
Beauvericin	BeSYN	352	*B. bassiana*	*S. cerevisiae*	Plasmid pXW55	Yu, 2013, 2017
Bassianolide	BaSYN	348	*B. bassiana*	*S. cerevisiae*	Plasmid pXW55	Yu, 2013, 2017, 2019
Brevianamide F	FtmA	319*	*A. fumigatus*	*S. cerevisiae*	Plasmid pKW1250	Tsunema-tsu, 2013
Acetylaszonalenin	AnaPS	280	*N. fischeri*	*S. cerevisiae*	Plasmid pXW55	Gao, 2012
Asperlicin C/ D	AspA	276	*A. alliaceus*	*S. cerevisiae*	Plasmid pXW55	Gao, 2013
Indigoidine	BpsA	141	*S. lavendulae*	*S. cerevisiae*	Genomic integration	Wehrs, 2018
$_{D}$Phe-$_{L}$Leu	TycA SrfAC	124 144	*B. parabrevis* *B. subtilis*	*S. cerevisiae*	Plasmid pESC	Siewers, 2010
2,4-dihydroxy 5,6-dimethyl benzaldehyde	ATEG_03630	120	*A. terreus*	*S. cerevisiae*	Plasmid pRS414	Wang, 2014
Phenguignardic acid	PgnA	104*	*A. terreus*	*S. cerevisiae*	Plasmid pESC	Huhner, 2017
Atromentin	ATEG_03090	101*	*A. terreus*	*S. cerevisiae*	Plasmid pESC	Huhner, 2017
Aspulyinone E	Apva & MelA	92*	*A. terreus*	*S. cerevisiae*	Plasmid pESC	Huhner, 2017

1974 *Estimated on the basis of the amino acid sequence length (with an average of 110 Da per amino acids)

1975 **Table 3:** Previous studies reporting NRP heterologous production in filamentous fungi

Natural product (NRP)	NRPS	NRPS MW (kDa)	Native producer	Heterologous host	Expression system	Reference
KK-1	KK-1 NRPS	1,430*	*C. clavata*	*A. oryzae*	Genomic integration	Yoshimi, 2018
Cycloaspeptide A/E	PscyB	623*	*P. sopii*	*A. oryzae*	Plasmid pTYGS-arg	Mattos-Shippley, 2018
Ferrirhodin	FSN1	495*	*F. sacchari*	*A, oryzae*	Plasmid pTAYAGSarg	Munawar, 2013
ACV	ACVS (or PcbAB)	450	*P. chrysogenum*	*N. crassa* *A. niger*	Cosmid Pcx3.2	Smith, 1990
				A. nidulans	Plasmid pJET1.2	Unkles, 2014
Bassianolide	BaSYN	350	*B. bassiana*	*A. niger*	Genomic integration	Boecker, 2018 Steiniger, 2017
CDP hexa-bassianolide	BaSYN-EnSYN TC_3	350	*B. bassiana* *F. oxysporum*	*A. niger*	Genomic integration	Steiniger, 2017
Beauvericin	BeSYN	350	*B. bassiana*	*A. niger*	Genomic integration	Boecker, 2018 Steiniger, 2017
Octa-beauvericin	BeSYN-BaSYN TC_3	350	*B. bassiana*	*A. niger*	Genomic integration	Steiniger, 2017
[PheLac]-Beauvericin	PSYN M_1-BeSYN M_2	350	*R. abscondita* *B. bassiana*	*A. niger*	Plasmid pVG2.2	Zobel, 2016
Enniatin	EnSYN	350	*F. oxysporum*	*A. niger*	Genomic integration	Richter, 2014 Steiniger, 2017
Octa-enniatin	EnSYNΔC_3-BaSYN C_3	350	*F. oxysporum* *B. bassiana*	*A. niger*	Genomic integration	Steiniger, 2017
[PheLac]-Enniatin	PSYN M_1-EnSYN M_2	350	*R. abscondita* *F. oxysporum*	*A. niger*	Plasmid pVG2.2	Zobel, 2016
Brevianamide F	FtmA	319*	*A. fumigatus*	*A. nidulans*	Genomic integration	Maiya, 2006
				A. niger	Plasmid pKW20142	Tsunematsu, 2013
			N. fischeri	*A. nidulans*	Genomic integration	Wunsch, 2015
Basidioferrin	CsNPS2	275	*C. subvermispora*	*A. niger*	Genomic integration	Brandenburger, 2017
Asperphenamate	ApmA ApmB	275* 275*	*P. brevicompactum*	*A. nidulans*	Genomic integration	Li, 2018
Gliotoxin	GliP	236	*A. fumigatus*	*A. nidulans*	Genomic integration	Chang, 2013
Microperfuranone	MicA	103*	*A. nidulans*	*A. niger*	Plasmid pSM565	Yeh, 2012
Aspulvinone E	ApvA MelA	92*	*A. terreus*	*A. nidulans*	Genomic integration	Guo, 2015
				A. niger		Geib, 2016, 2017

1976 *Estimated on the basis of the amino acid sequence length (with an average of 110 Da per amino acids)

Figure legends

Figure 1: Multiple steps involved in NRP biosynthesis pathway. **A**- Transcription of the NRPS encoding gene. **B**- Translation of the NRPS mRNA. **C**- Folding of the NRPS in multi-enzymatic domains modular organization and addition of phosphopantetheine arm (*holo*-NRPS). Several domains are presented: A: adenylation; T: thiolation; C: condensation; C_{term}/Te: thioesterase or terminal condensation domain (only found in fungal NRPSs); and an example of a potential additional epimerization (E) domain (converting an $_{L}$-form amino acid to its $_{D}$-form). **D**- Operating way of NRPS assembly line to generate NRP (Mm: monomer). *Epimerization reaction can occur either before or after condensation reaction.

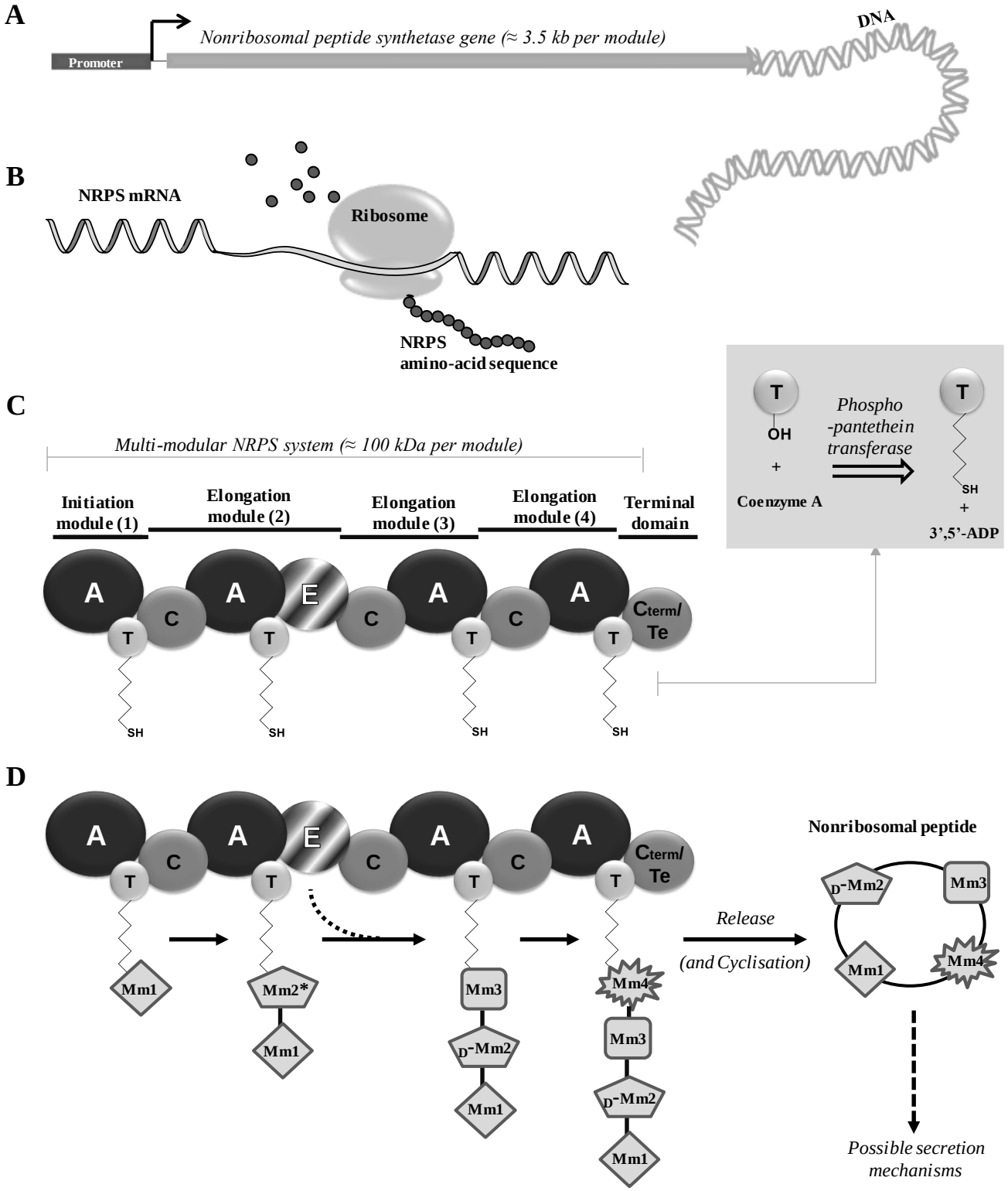
Figure 2: Simplified workflow for novel NRP discovery by genome mining with an overview of the main bioinformatics tools available (accessible through hyperlinks).

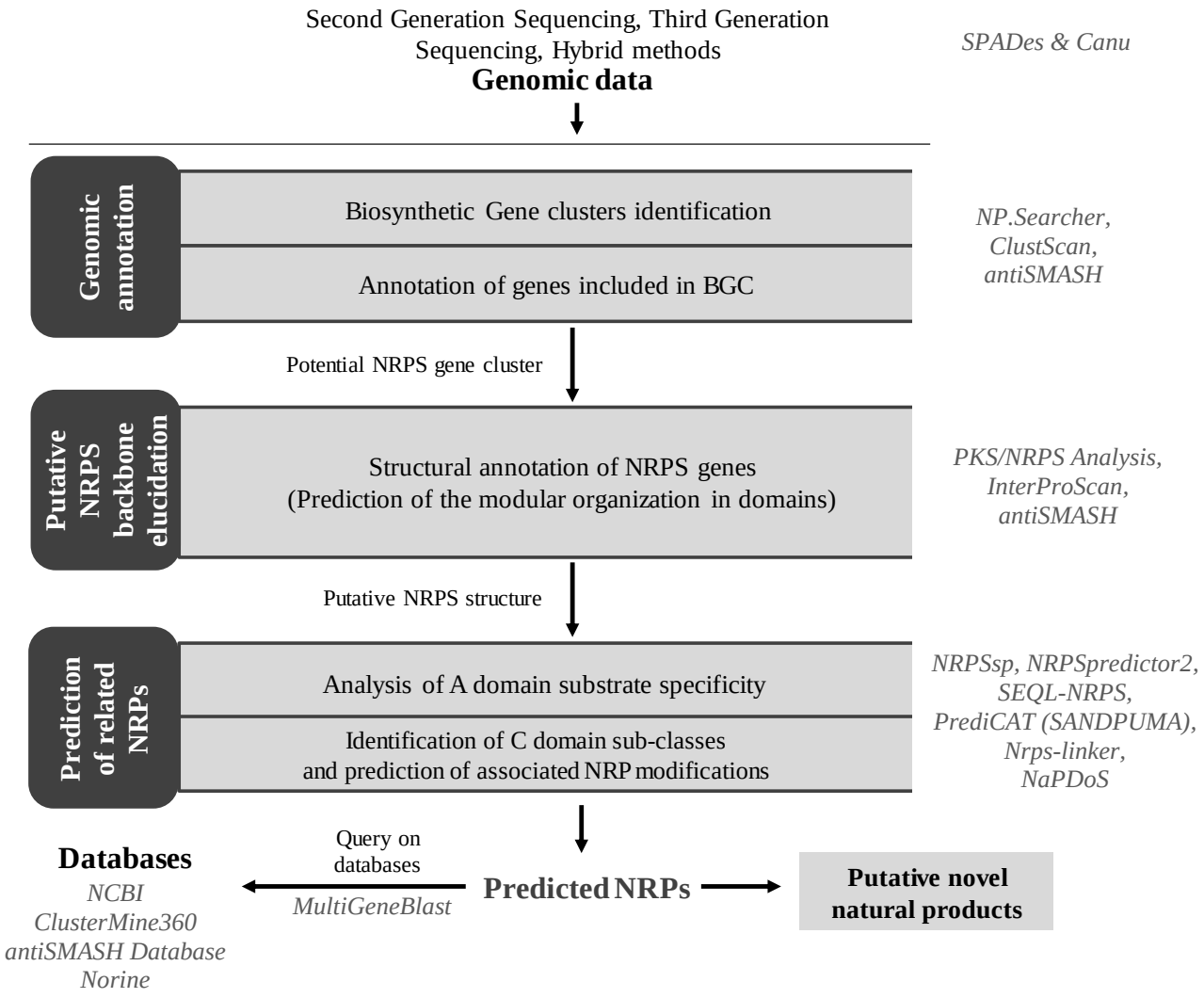
Figure 3: Analysis of the 50 most frequent monomers constituting the nonribosomal peptides classified in the Norine database. Each bar is subdivided into colored parts corresponding to the taxonomic clade from which the resulting NRPs was identified (i.e., white: bacteria, dark grey: fungi). The monomers specifically activated by A-domains which display pHMMs signatures enabling specificity predictions are highlighted in green.

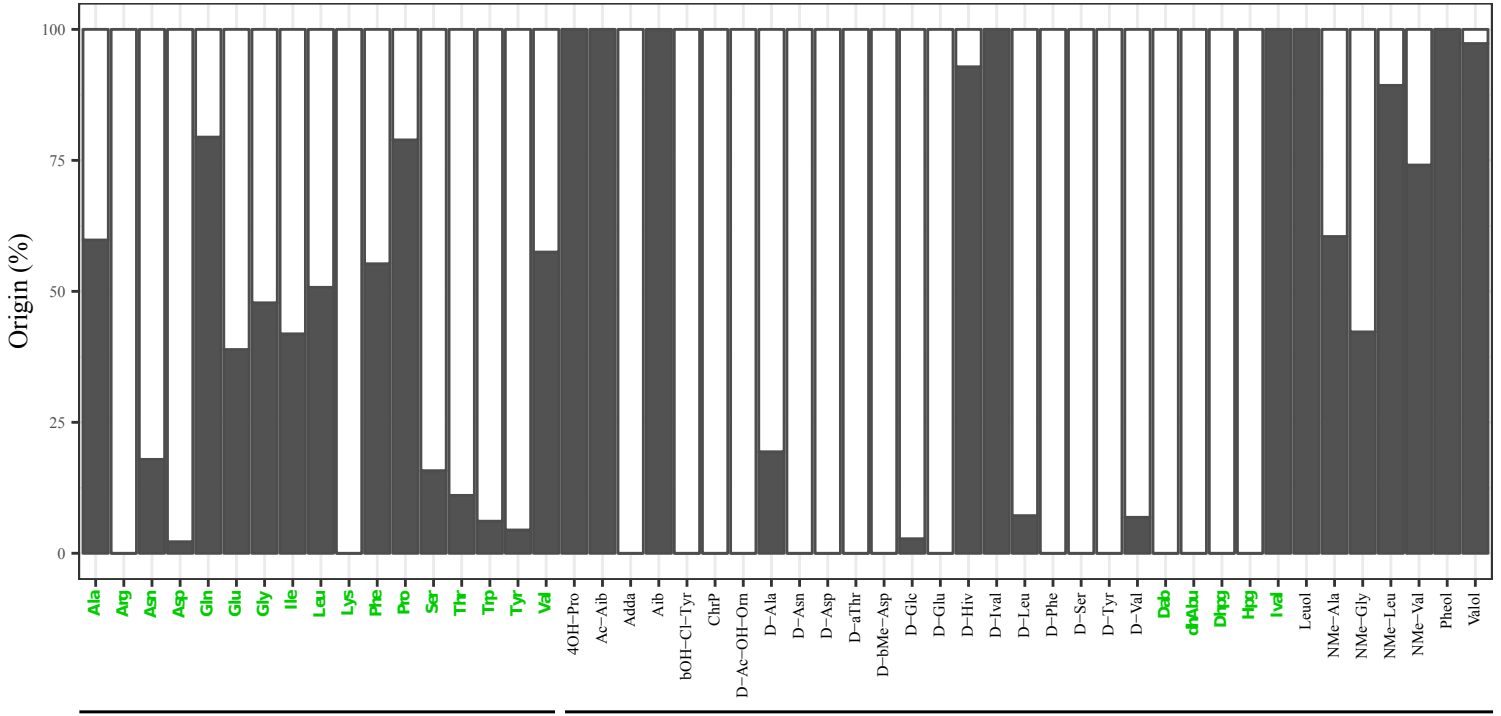
Figure 4: Description of the four main cloning methods enabling the reconstitution, from multiple amplified PCR fragments, of long-length NRPS sequences into a targeted vector: **A.** Overlap extension PCR approach, **B.** Gibson assembly, **C.** USER Cloning, and **D.** TAR Cloning (homologous recombination).

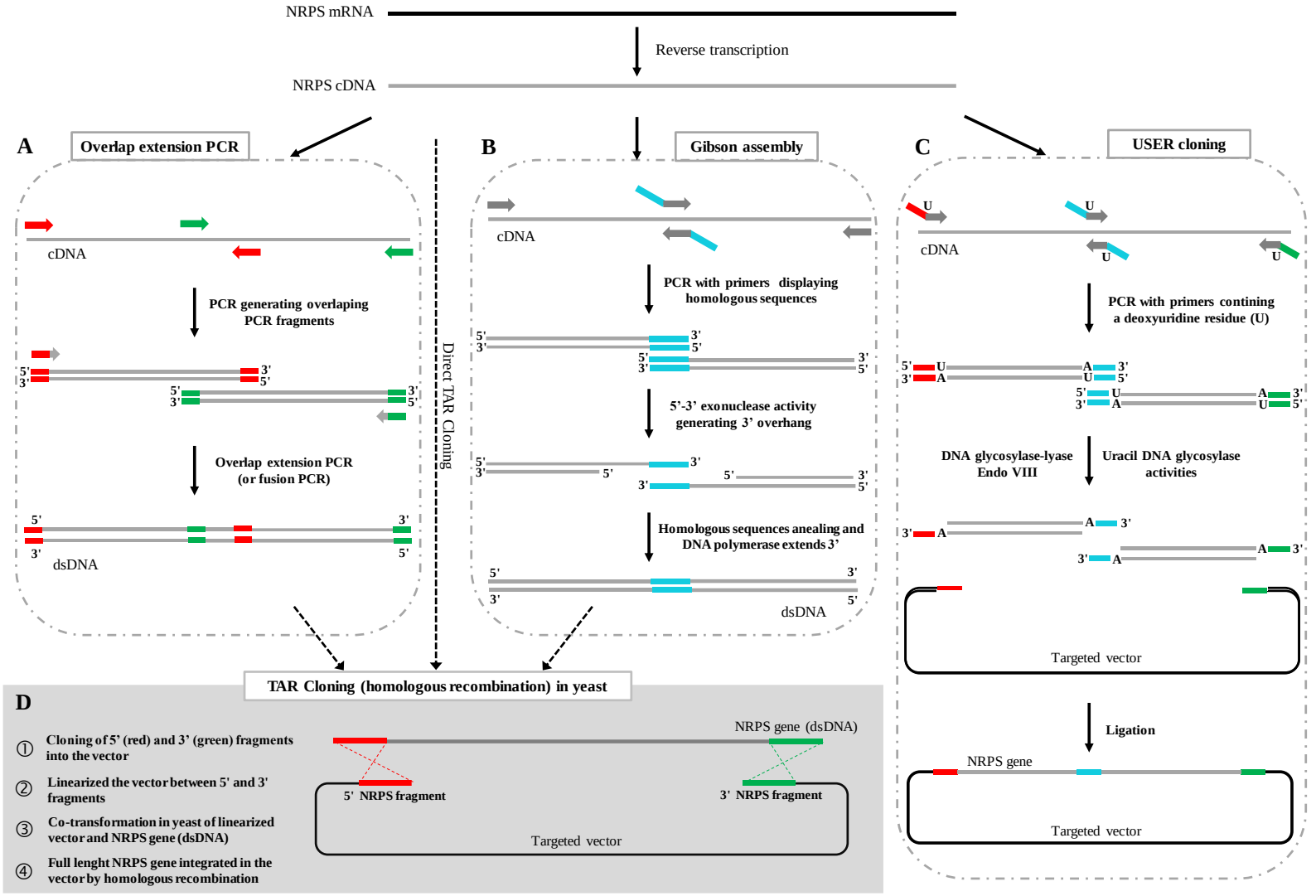
Figure 5: Overview of the limiting steps of NRP heterologous production and of the main solutions to overcome them and consider upgrading production rates.

Figure 6: A. Presentation of the gene cluster involved in the biosynthesis of penicillin in *Penicillium chrysogenum*. **B**. Schematic representation of the penicillin G biosynthetic pathway in *P. chrysogenum*. ACV is assembled by the ACVS (encoded by the *pcbAB* gene), before to be modified into isopenicillin N by IPNS (encoded by *pcbC* gene), subsequently converted into penicillin G through combined activities of IAT and PCL (respectively encoded by *penDE* and *pclA* gene). *PclA** gene is not part of the penicillin biosynthetic cluster.


A
Promoter
Nonribosomal peptide synthetase gene (≈ 3.5 kb per module)
DNA
B
NRPS mRNA
Ribosome
NRPS
amino-acid sequence
C
Multi-modular NRPS system (≈ 100 kDa per module)
Initiation module (1)
Elongation module (2)
Elongation module (3)
Elongation module (4)
Terminal domain
A
C
T
E
Cterm/ Te
SH
T
OH
+
Coenzyme A
Phospho -pantethein transferase
SH
+
3',5'-ADP
D
Mm1
Mm2*
Mm3
D-Mm2
Mm4
Release
(and Cyclisation)
Nonribosomal peptide
Possible secretion mechanisms


Second Generation Sequencing, Third Generation Sequencing, Hybrid methods
Genomic data
SPADes & Canu
Genomic annotation
Biosynthetic Gene clusters identification
Annotation of genes included in BGC
NP.Searcher, ClustScan, antiSMASH
Potential NRPS gene cluster
Putative NRPS backbone elucidation
Structural annotation of NRPS genes (Prediction of the modular organization in domains)
PKS/NRPS Analysis, InterProScan, antiSMASH
Putative NRPS structure
Prediction of related NRPs
Analysis of A domain substrate specificity
Identification of C domain sub-classes and prediction of associated NRP modifications
NRPSsp, NRPSpredictor2, SEQL-NRPS, PrediCAT (SANDPUMA), Nrps-linker, NaPDoS
Databases
NCBI
ClusterMine360
antiSMASH Database
Norine
Query on databases
MultiGeneBlast
Predicted NRPs
Putative novel natural products

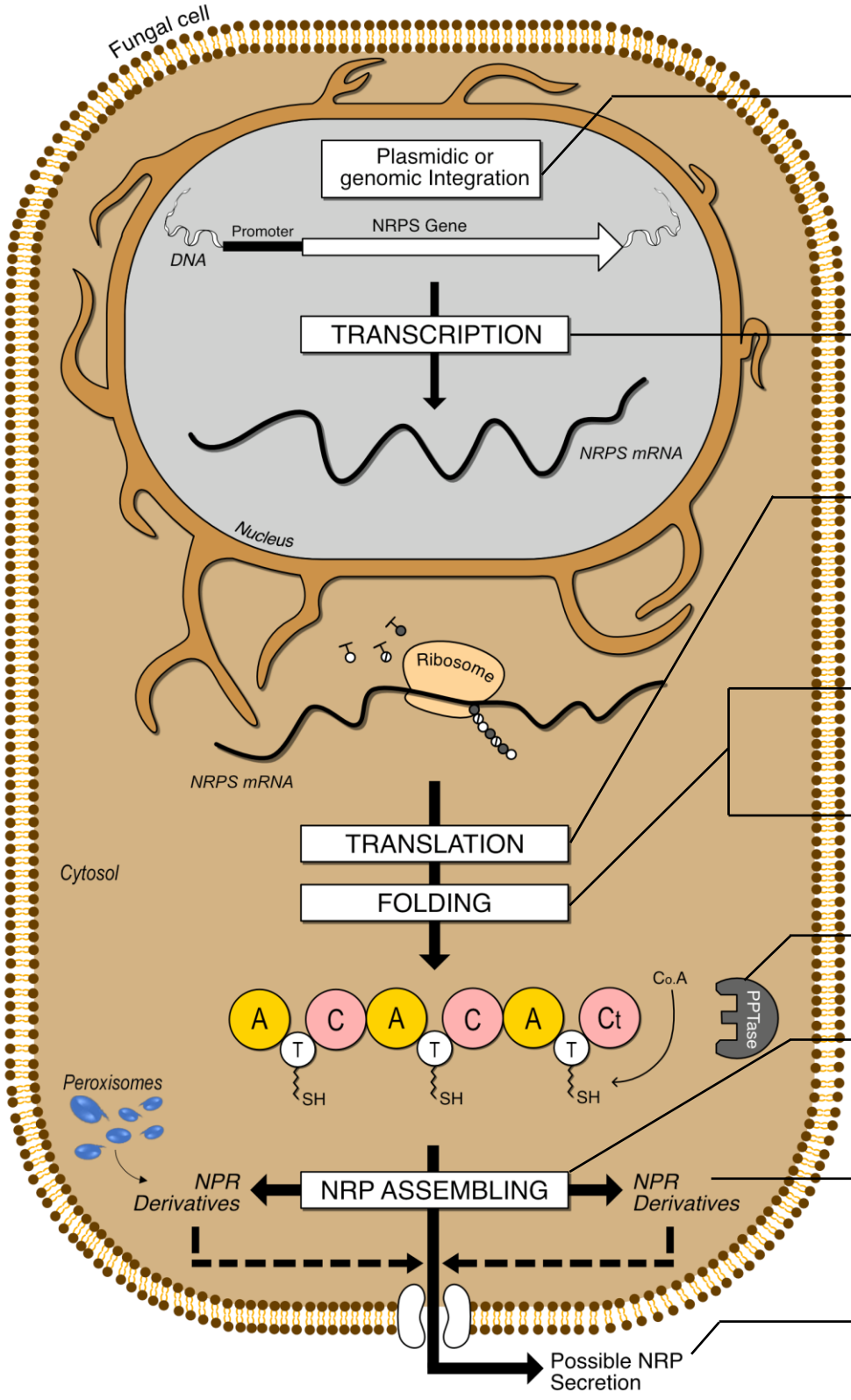

Origin (%)
100
75
50
25
0
Ala
Arg
Asn
Asp
Gln
Glu
Gly
Ile
Leu
Lys
Phe
Pro
Ser
Thr
Trp
Tyr
Val
4OH−Pro
Ac−Aib
Adda
Aib
bOH−Cl−Tyr
ChrP
D−Ac−OH−Orn
D−Ala
D−Asn
D−Asp
D−aThr
D−bMe−Asp
D−Glc
D−Glu
D−Hiv
D−Ival
D−Leu
D−Phe
D−Ser
D−Tyr
D−Val
Dab
dHAbu
Dhpg
Hpg
Ival
Leuol
NMe−Ala
NMe−Gly
NMe−Leu
NMe−Val
Pheol
Valol
Proteinogenic amino acids
Exotic and D amino acids


NRPS mRNA
Reverse transcription
NRPS cDNA
A
Overlap extension PCR
cDNA
PCR generating overlaping PCR fragments
Overlap extension PCR (or fusion PCR)
dsDNA
Direct TAR Cloning
B
Gibson assembly
cDNA
PCR with primers displaying homologous sequences
5'-3' exonuclease activity generating 3' overhang
Homologous sequences anealing and DNA polymerase extends 3'
dsDNA
C
USER cloning
cDNA
PCR with primers contining a deoxyuridine residue (U)
DNA glycosylase-lyase Endo VIII
Uracil DNA glycosylase activities
Targeted vector
Ligation
NRPS gene
Targeted vector
TAR Cloning (homologous recombination) in yeast
D
① Cloning of 5' (red) and 3' (green) fragments into the vector
② Linearized the vector between 5' and 3' fragments
③ Co-transformation in yeast of linearized vector and NRPS gene (dsDNA)
④ Full lenght NRPS gene integrated in the vector by homologous recombination
NRPS gene (dsDNA)
5' NRPS fragment
3' NRPS fragment
Targeted vector

Limiting steps of NRP heterologous production

[Corresponding text section]

❖Main solutions to overcome them



Issues related to long-lenght NRPS gene integration *[3.2.1]*

❖ Use of dedicated cloning methods
- TAR Cloning
- PCR overlap extension
- USER Cloning
- Gibson assembly

Very low or inexistent transcription *[3.3][4]*

❖ Modification of the promoter
❖ Modulation of the promoter strenght
❖ Addition of transcription factors

Very low or inexistent translation *[3.2.1][3.4.1]*

❖ Elimination of potentially mis-spliced introns (with cDNA cloning at DNA level)
❖ Codon optimization (at DNA level)
❖ Cloning of genes encoding missing tRNA

NRPS misfolding *[3.4.1]*

❖ Implement foreign chaperone proteins
❖ Cultivation at lower temperature

Degradation of the NRPS structure *[3.4.1]*

❖ Select host strain mutated in protease encoding genes

Inactivated (Apo-) NRPS *[3.4.2]*

❖ Cloning of heterologous PPTase gene

NRP precursor deficiency *[3.4.3]*

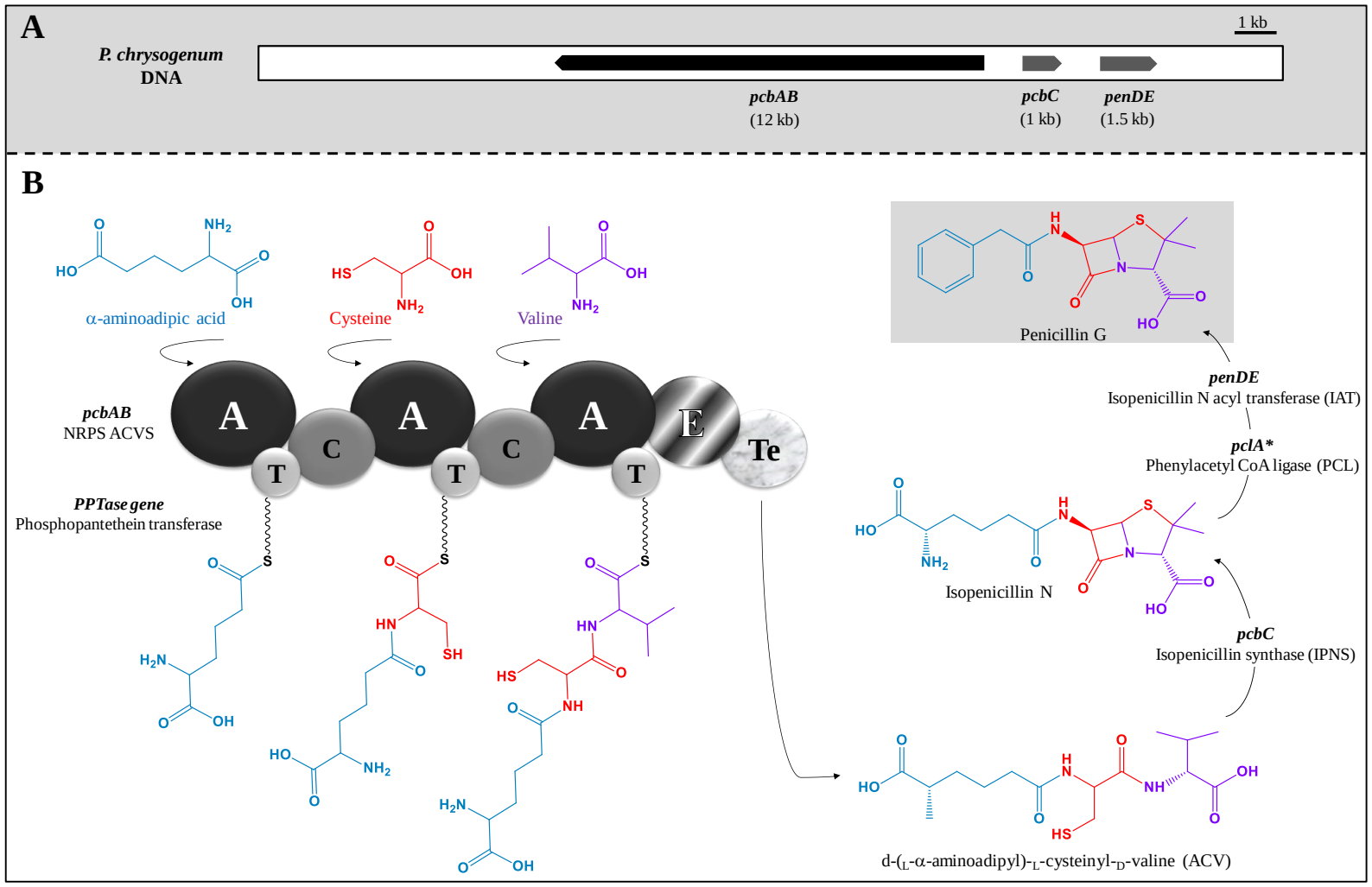
❖ Feeding with precursors
❖ Cloning of genes involved in precursor biosynthesis

Absence of NRP derivatives *[3.4.2]*

❖ Cloning of missing tailoring enzymes
❖Addition of a targeting signal to address enzymes in a cellular compartment

Potential NRP toxicity *[3.4.4]*

❖ Implement a secretion system
❖ Cloning of genes confering a resistance
❖ Extrude NRP in organic solvents


A
1 kb
P. chrysogenum
DNA
pcbAB
(12 kb)
pcbC
(1 kb)
penDE
(1.5 kb)
B
α-aminoadipic acid
Cysteine
Valine
pcbAB
NRPS ACVS
A
C
A
C
A
E
Te
T
T
T
PPTase gene
Phosphopantethein transferase
Penicillin G
penDE
Isopenicillin N acyl transferase (IAT)
pclA*
Phenylacetyl CoA ligase (PCL)
Isopenicillin N
pcbC
Isopenicillin synthase (IPNS)
d-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV)