Expression of fungal biosynthetic gene clusters in *S. cerevisiae* for natural product discovery

Zihe Liu, Zhenquan Lin, Jens Nielsen

**A B S T R A C T**

Fungi are well known for production of antibiotics and other bioactive secondary metabolites, that can be served as pharmaceuticals, therapeutic agents and industrially useful compounds. However, compared with the characterization of prokaryotic biosynthetic gene clusters (BGCs), less attention has been paid to evaluate fungal BGCs. This is partially because heterologous expression of eukaryotic gene constructs often requires replacement of original promoters and terminators, as well as removal of intron sequences, and this substantially slow down the workflow in natural product discovery. It is therefore of interest to investigate the possibility and effectiveness of heterologous expression and library screening of intact BGCs without refactoring in industrial friendly microbial cell factories, such as the yeast *Saccharomyces cerevisiae*. Here, we discuss the importance of developing new research directions on library screening of fungal BGCs in yeast without refactoring, followed by outlook prominent opportunities and challenges for future advancement.

1. Introduction

Microorganisms provide us with a large number of biosynthetic gene clusters (BGCs) that produce bioactive secondary metabolites [1]. However, many of these are being produced by microorganisms that grow slowly, or are even unculturable at laboratory conditions [2]. Moreover, the majority of secondary metabolites are with complex structures, and are expressed at low levels, or even silent at most conditions [3]. Thus, extraction or chemical synthesis of natural products is often costly, and bioproduction through industrial friendly hosts is gaining increased attentions [4].

Recent microbial genome sequences studies showed that awaken the silent or cryptic BGCs would enlarge the reservoir of natural products and provide more opportunities for identifications of novel compounds [5,6]. However, the number of identified natural products is far less compared with the number of putative BGCs [7], and the low-throughput expression step, either by endogenous activation [8,9] or heterologous refactoring [10,11], is one of the rate limiting steps in the pipeline of natural product discovery. For example, efforts on activation of BGCs in native hosts have been successful in a variety of cases, however, these advancements will not be helpful for the 99% of microbial strains that are not readily cultivable at laboratory conditions [2]. Moreover, compared with prokaryotic BGCs, heterologous expression for eukaryotic BGCs requires refactoring each coding region of the whole gene clusters, which adds additional challenges.

Fungi produce a wide range of natural products. For example, many compounds from fungi possess antimicrobial activities such as the beta-lactam antibiotics penicillin and cephalosporins, as well as the antifungal griseofulvin [12]. However, compared with those from bacteria, less attentions have been paid to investigate fungal BGCs [13]. This is partially because heterologous refactoring of eukaryotic BGCs often requires replacement with native promoters and terminators, as well as removal of intron sequences [14]. Thus, it will substantially speed up the pipeline of natural product discovery if we could develop engineered yeast cell factories having improved intron splicing and recognition of promoters from target fungi.
2. Key research directions

In order to achieve high-throughput library expression of heterologous BGCs in a cut and paste fashion without refactoring (Fig. 1), there are at least three research directions that need to be optimized in S. cerevisiae (Fig. 2): 1) optimization of fungal BGC expression, 2) identification of target BGCs and 3) development of high-throughput BGC cloning. Powerful analytical platforms to identify novel natural products is also of great importance, but will not be covered further here as it is extensively covered in recent reviews [15–17].

2.1. Genetic elements and natural product discovery

Heterologous expression of eukaryotic BGCs usually requires removal or replacement of original transcription and translation regulation elements, such as introns, promoters, and terminators (Fig. 2). The spliceosome is host specific, and several widely used microbial cell factories such as S. cerevisiae generally cannot remove introns from distant fungal species [10,14,18]. This problem could be tackled by directly generating cDNAs of target BGCs [19], however, this could only be applied to expressed BGCs within the ~1% cultivable species [14]. For silent BGCs or the BGCs from the environment samples, another solution is in vitro assembly exons of a given gene or chemical synthesis. However, current in silico intron prediction tools have limitations, and mis-annotated exons, even with 1bp variation, can cause impaired protein translations and ruin the effort of the whole BGC characterization [20]. On the other hand, since a large number of fungal BGCs contain more than ten genes, it is challenging to replace promoters and terminators for each gene to generate expression cassettes, not to mention the high-throughput cloning. Thus, in order to multiplex expression of BGCs in yeast without refactoring and screen for novel natural products, it is important to engineer and evolve yeast to recognize heterologous transcription units or BGCs.

2.2. Bioinformatic analysis and natural product discovery

Recent development in sequencing technologies and bioinformatic tools have greatly advanced BGC discovery. We now have tremendous genome sequences available on line for natural product discovery (Fig. 2). Databases and annotation tools are being developed and successfully applied for BGC annotation and natural product discovery. For example, we can use Bayesian statistics to perform phylogenetic analysis [23]; eukaryotic orthologous groups (KOGs) [24] and Kyoto Encyclopedia of Genes and Genome (KEGG, http://www.kegg.jp/) for annotation analysis; antiSMASH [25], BiG-SCAPE [26], SMURF [27] and Pseudomonas sp. for chemical identification; Screening in S. cerevisiae and Multiplexed BGCs cloning (Fig. 2). A number of cloning techniques have been reported for heterologous expression of target BGCs (Fig. 2). These approaches could be divided into sequence-independent library cloning methods that screen natural products from random sequenced genomes, and direct cloning methods that identify novel natural products based on precise bioinformatic annotations followed by cloning target BGCs in a low throughput fashion [32]. Current heterologous BGC cloning techniques have achieved vast progress [32,33], however, the process is usually laborious and time consuming. For example, regarding the sequence-independent library cloning method, in order to cover reasonable number of BGCs, 10-20 fold-coverage of the whole genome or metagenomes needs to be generated [34]. Moreover, current direct cloning methods need to replace or remove all heterologous transcription and translation regulation elements, e.g. promoters, terminators, introns, and have been limited to capture few clusters per round [10,11,18,19]. With CRISPR tools revolutionizing the field of genome editing, it would be interesting to combine the strength of both methods, and simultaneously clone and screen all putative BGCs of a given genome without refactoring. We anticipate that this could work through optimization of high-throughput capture and cloning techniques. The step of enrichment of target BGCs after CRISPR-assisted in vitro cleavage may also need to be optimized.

3. Summary

The ability of heterologous expression and library screening of intact fungal BGCs without refactoring will contribute substantially to the field of synthetic biology and natural product discovery. In order to fulfill the great demand of natural products, it is necessary to develop a range of novel technologies that can speed up the fungal natural products discovery pipeline. For example, as natural products are derived from a limited number of precursor metabolites and co-factors [1], such as short chain carboxylic acids, amino acids, NADH, we need to engineer the primary metabolism to ensure efficient provision of precursor supplies. We may also need to improve the yeast capability of transcription and post-transcriptional-modification of fungal BGCs, with focus primarily on promoter recognition and intron splicing. We may also need to develop CRISPR-based high-throughput and multiplexed BGC cloning.

Fig. 1. Multiplexed screening of Penicillium BGCs in yeast. Putative BGCs can be analyzed by well-developed bioinformatics tools, and then captured through direct cloning for heterologous expression and characterization in S. cerevisiae.

Fig. 2. Workflow of high-throughput fungal natural product discovery in yeast. Development of strategies of multiplexed expression of fungal BGCs in yeast could be divided optimization of BGC expression, identification of target BGCs and development of high-through BGC cloning.

MIBiG [28] to identify putative BGCs; Softberry (http://www.softberry.com) to predict intron sequences; Mauve [29] to perform comparative genome analysis, etc. Moreover, we could further narrow down target BGCs through literature research, MIBiG database and FungiFun [30] to analyze key enzymes of selected BGCs. So far, millions putative BGCs have been predicted [31], however, the low throughput cloning and analytic techniques has substantially slowed down the discovery novel natural products and associated BGCs [7].
Acknowledgment

The authors indicate that they have no conflict of interest.

Zhihe Liu: Writing - original draft, drafted the outline and wrote the manuscript. Zhenquan Lin: Writing - original draft, drafted the outline and wrote the manuscript. Jens Nielsen: Writing - original draft, drafted the outline and wrote the manuscript. Supervision, supervised the research. All authors have read and approved the final manuscript.

Declaration of competing interest

The authors indicate that they have no conflict of interest.

References