Published in final edited form as:

Curr Top Med Chem. 2016 ; 16(15): 1695–1704.

Direct Capture Technologies for Genomics-Guided Discovery of Natural Products

Andrew N. Chan[1], **Kevin C. Santa Maria**[1], and **Bo Li**[1,*]

[1]Department of Chemistry, Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Abstract

Microbes are important producers of natural products, which have played key roles in understanding biology and treating disease. However, the full potential of microbes to produce natural products has yet to be realized; the overwhelming majority of natural product gene clusters encoded in microbial genomes remain “cryptic”, and have not been expressed or characterized. In contrast to the fast-growing number of genomic sequences and bioinformatic tools, methods to connect these genes to natural product molecules are still limited, creating a bottleneck in genome-mining efforts to discover novel natural products. Here we review developing technologies that leverage the power of homologous recombination to directly capture natural product gene clusters and express them in model hosts for isolation and structural characterization. Although direct capture is still in its early stages of development, it has been successfully utilized in several different classes of natural products. These early successes will be reviewed, and the methods will be compared and contrasted with existing traditional technologies. Lastly, we will discuss the opportunities for the development of direct capture in other organisms, and possibilities to integrate direct capture with emerging genome-editing techniques to accelerate future study of natural products.

Keywords

Direct capture; ET recombination; Genomics; Homologous recombination; Lambda red; Natural products; Transformation-associated recombination

1. INTRODUCTION

Natural products have revolutionized modern medicine. Between the 1940s and 1960s, natural products of diverse structures and biological activities were frequently isolated from microbial cultures and transformed into life-saving pharmaceuticals, including antibiotics, anticancer agents, and immunosuppressants. However, discovery efforts were greatly reduced towards the end of the 20th century because the same most abundant natural products were being “rediscovered” in bioactivity-based screens. As a result, natural product

*Address correspondence to at the Department of Chemistry, University of North Carolina at Chapel Hill, CB 3290, Chapel Hill, NC, 27599, USA; Tel/Fax: +1-919-843-4436; boli@email.unc.edu.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflict of interest.

discovery efforts experienced largely diminishing returns. In the post-genomic era, natural product research has not only gained renewed interest, but has also undergone dramatic transformations propelled by advanced sequencing technologies and genomic tools. Genomics has revealed the potential of microbes to produce more natural products than previously thought, providing a method to circumvent the problems of rediscovery. The workflow of genomics-guided discovery is depicted in Figure 1 using *Streptomyces clavuligerus* as an example, the genome of which contains more than 40 putative biosynthetic gene clusters [1].

Genomics-guided discovery of natural products is likely to have profound impacts in multiple fields. While natural products remain among the most promising drug leads for many contemporary diseases, their importance and applications are becoming increasingly recognized in the fields of synthetic biology, chemical ecology and microbiology. Enzymes involved in secondary metabolism catalyze a diverse set of reactions that can be evolved and utilized in synthetic biology. Natural products themselves play important roles in mediating microbe-microbe interactions, host-microbe interactions and influencing disease, growth and development. Structural and mode of action studies of these compounds hold great promise for advancing our understanding of these processes.

While elucidating the genes involved in the biosynthesis of these molecules is a daunting task, it is aided by the tendency of natural product biosynthetic genes to “cluster” within the genome. Genes involved in the synthesis of a single natural product are generally found in the same genetic locus. Multiple challenges need to be overcome to link these clusters to chemical compounds. First, many “orphan” gene clusters either do not express in laboratory conditions, or at levels too low for product detection. Second, many natural products come from bacteria that are difficult to grow or manipulate, including numerous “unculturable” microbes from the human microbiome and environmental sources. Third, natural product gene clusters can reach over 100 kb in size, especially those that involve the assembly line-like non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs), further increasing the difficulty of genetic manipulation. Finally, identification and structural characterization remains a low-throughput effort that requires both specialized skills and intensive efforts.

The development of new technologies that allow for the facile and efficient connection of genetic information to secondary metabolites is crucial to modern discovery efforts. As a result, a number of tools have been developed and implemented to activate orphan gene clusters, such as mutagenesis, modification of regulatory elements, ribosome engineering, stimulation with environmental factors, and interspecies interactions [4]. These methods have led to the discovery of a number of new natural products and the readers are referred to several of excellent reviews for more details [4–8]. In this review, rather than providing a comprehensive review of tools for natural product discovery, we have chosen to focus on recent developments in direct capture technologies for the heterologous production of natural products. These technologies take advantage of well-characterized genetic systems, bypass the need for culturing and manipulating native producers, and have potentially broad applications and high-throughput capacity.

2. OVERALL CONCEPT OF DIRECT CAPTURE

Direct capture utilizes homologous recombination to isolate gene clusters from genomic DNA in a single step for later heterologous expression. Homologous recombination based techniques have emerged as powerful tools for genetic manipulations. This strength is evidenced from both the yeast knockout collection and the *Escherichia coli* Keio collection, both generated by homologous recombination [9–10]. Direct capture methods utilize this approach to manipulate large DNA molecules in a single step. Two main strategies have emerged in direct capture. The first utilizes the model organism *Escherichia coli* and phage recombination (λ Red/ET Recombination) similar to that in the construction of the Keio collection. The second takes advantage of the endogenous homologous recombination machinery in *Saccharomyces cerevisiae* (transformation-associated recombination) as in the yeast knockout collection [9–10]. The principles and applications of both strategies will be discussed in this review.

In both techniques, a capture vector is designed with arms that are homologous to the boundaries of the target gene cluster. The “capture host,” either *E. coli* or *S. cerevisiae,* is then co-transformed with the linearized vector and donor DNA where the target gene cluster resides. Homologous recombination results in the capture of the target cluster in the vector, which can be readily transferred to a heterologous host for compound production (Fig. 2). In addition to identifying new natural products, direct capture establishes a direct link between DNA and small molecules. In a way that the central dogma of molecular biology connects genes to proteins, we can now connect sets of genes to compounds [11]. This knowledge and understanding will help to inform future discovery efforts.

3. METHODOLOGIES

3.1 Natural Product Gene Cluster Identification and Prediction

The first step of genomics-guided natural product discovery is to identify novel natural product gene clusters from genome sequences. This effort has been accelerated by recent developments in bioinformatics and automation. A number of software platforms, notably antiSMASH, Cluster-finder, and SMURF, have been developed to identify genes likely involved in secondary metabolism using Hidden Markov Models [2, 12–13]. Some of these searches take advantage of the modular nature of PKS and NRPS assembly lines, as well as the individual biosynthetic domains within, to predict structural elements of the gene cluster’s product [14–17]. Such *in silico* structure predictions have enabled the recent identification of the novel cyclic lipopeptide orfamide A, as well as the pentapeptide precursor to nocardicin A [18–20].

Determining the boundaries of the cluster remains less straightforward than structural prediction of the core molecule. If the predicted boundaries are too narrow, important biosynthetic genes may be left out. Conversely, if the boundary predictions are too wide, in addition to increasing the difficulty of understanding the pathway, unrelated clusters may be grouped together as “hybrid” clusters. In an effort to combat this, many bioinformatics genome mining platforms use a training set to optimize cluster boundary predictions. When assigning genes to either the same or separate clusters, they may take into account the

number of non-secondary metabolism genes between potential boundaries, the total number of genes, and the number of nucleotides between genes [18–20]. Manual inspection of cluster boundaries still has utility, especially when the cluster is present in multiple sequenced organisms. The synteny between related clusters can aid in identifying the genes that are part of the cluster, and those that are unrelated [21].

Even when cluster boundary predictions are accurately made, a number of cases exist in which natural product gene clusters do not follow the simple ‘rule’ of clustering and instead involve additional genes at distal loci. In the biosynthesis of actinorhodin by *Streptomyces coelicolor* A3(2), a malonyltransferase essential in primary metabolism is necessary for biosynthesis, and is located 2.8 Mb from the actinorhodin gene cluster [22]. A similar situation is seen in the biosynthesis of the histone deacetylase inhibitors FK228 and the thailandepsins [23–24]. In all three of these clusters, the analogous gene for secondary metabolism was absent, and successful compound production required either the identification of the responsible gene from elsewhere in the genome, or the utilization of a heterologous host capable of supplying its own analogous gene. In the case of the siderophores erythrochelin and rhodochelin, two and three distinct gene clusters are needed, respectively, for biosynthesis of the natural product. The phytotoxin coronatine is composed of two separate moieties, coronafacic acid and coronamic acid, and in many strains of *Pseudomonas syringae*, the genes for both are within a single cluster. However, in *P. syringae* pv. *tomato* DC3000, the genes for each moiety are separated by 26 kb [25]. In a similar case, while the biosynthesis of the antibiotic congocidine is limited to a single gene cluster in *Streptomyces ambofaciens*, its biosynthesis in *Streptomyces netropsis*, as well as that of two other pyrrolamides, requires two gene clusters separated by at least 58 kb [26]. These cases are only a small subset of known natural product bio-synthetic pathways, but illustrate the concept that even the best gene cluster prediction software may not reveal all essential genes for a given product.

Ultimately, there is no single correct method to accurately identify the boundaries of a gene cluster. However, the power of bioinformatics platforms and the number of microbial genomes available have greatly simplified the process [27]. Based on the examples described later in this review, the approach of including extra, unrelated genes within the predicted bounds of the cluster has allowed the isolation of a number of compounds, with no observed impacts (neither positive nor negative) on their production.

3.2.1. “Recombineering” in E. coli: λ Red/ET Recombination—Two parallel systems have been developed from phage recombination for use in *E. coli* as recombineering techniques. One system is derived from λ prophage, comprising the enzymes Redα, Redβ, and Redγ and is known as the “λ Red” system [28]. The analogous system from Rac prophage involves the enzymes RecE and RecT, known as “ET recombination” [29–30]. The pairs RecET and Redαβ are functionally equivalent. RecE and Redα are 5′ to 3′ exonucleases. RecT and Redβ are ssDNA-binding proteins, which facilitate complementary DNA strand annealing [31]. Redγ on the other hand, has no counterpart in the RecET system. The function of Redγ is to inhibit the activities of RecBCD, which function as dsDNA and ssDNA exonucleases and helicases [28, 32–33].

Both linear and circular DNA molecules have been successfully recombined with introduced linear DNA by phage recombination. The λ Red system works efficiently on circular substrates in a linear plus circular homologous recombination (LCHR) fashion. In contrast, the ET recombination system shows proficiency at linear plus linear homologous recombination (LLHR) [34]. It should be noted that the introduction of Redγ to ET recombination enhances the system, presumably due to protection of the introduced DNA from RecBCD's nuclease activities [32, 34]. Studies on LCHR and LLHR have revealed that they are mechanistically distinct. Essentially, LCHR is DNA replication dependent while LLHR was confirmed to be replication independent [34–35]. The details on the mechanism of phage recombination have been recently reviewed [32, 35–36].

The application of ET recombination in direct capture was first described in 2000, involving the capture of a large stretch of DNA (> 20 kb) from a BAC into a linearized plasmid (LLHR) [30]. In the same study, the capture of smaller targets (<5 kb) was also reported, by transforming an ET recombination proficient *E. coli* strain with genomic DNA fragments, in a linear plus linear fashion [30]. Fast-forward a decade, ET recombination was used in the capture of the 19 kb syringolin gene cluster from genomic DNA of *Pseudomonas syringae* (Fig. 2) [37]. The direct capture of this cluster from *P. syringae* allowed for its heterologous expression in *E. coli* and subsequent characterization of additional members of the syringolins, a family of hybrid PKS-NRPS natural products **(1)** (Fig. 3) [37]. In another study, full length RecE and RecT were used in the capture of the hybrid PKS-NRPS luminmycin cluster **(2)** and the NRPS luminmide cluster resulting in their heterologous production and characterization [34].

The λ Red system was also employed to successfully capture 20 natural product gene clusters and heterologously express them in engineered strains of *Streptomyces avermitilis* [38]. Although the λ Red system is less efficient than ET recombination for linear plus linear applications, it was successfully used in this study to capture gene clusters from linearized cosmids. The production of several clinically relevant and biologically important natural products was reported, including streptomycin (*Streptomyces griseus*), erythromycin A (*Sacchropolyspora erythraea*), cephamycin C (*Streptomyces clavuligerus*), holomycin **(3)** (*S. clavuligerus*), clavulanic acid **(4)** (*S. clavuligerus*), rebeccamycin (*Lechevalieria aerocolonigenes*), novobiocin (*Streptomyces anulatus*) and chloramphenicol (*Streptomyces venezuelae*) [38]. For many compounds, production was enhanced in comparison to the native producer. Streptomycin was improved relative to *S. griseus* by 2- to 2.5-fold. More dramatically, production of holomycin was increased from less than 0.2 mg/L in the wildtype producer to 8 mg/L in the heterologous expression strain [38–39]. In addition to improving production, the study was also able to activate a cryptic gene cluster. Although the *S. clavuligerus* genome encodes a gene cluster for pholipomycin, the strain has not been shown to produce such a compound. When this cluster was expressed in a derivative of *S. avermitilis*, a yield of 20 mg/L pholipomycin was observed [38]. This study demonstrated the versatility of the λ Red/ET recombination systems in capturing natural products of different classes and origins, while greatly enhancing production levels.

3.2.2. Expansion of Phage Recombination into Other Genera—Recently, the generation of a recombineering system in the related genera, *Photorhabdus* and

Xenorhabdus was reported [40]. Each harbors genes analogous to Red α, β, and γ, which have been repurposed for gene cluster activation by promoter exchange. The study introduced a tetracycline inducible promoter, using the analogous λ Red system, upstream of a cryptic chromosomal NRPS gene cluster to activate expression, resulting in the production of nonribosomal peptides ranging from a heptapeptide to a nonapeptide in the native bacterium [40]. Although this study did not involve direct capture of gene clusters from genomic DNA, it expanded on the phage homologous recombination system with potentially broad implications. The identification and functionalization of λ Red/ET recombination systems in other bacterial strains, especially those of the *Streptomyces* genus, may allow for the activation of cryptic gene clusters in their native hosts and bypass the need for heterologous hosts.

3.3.1. Yeast Transformation-Associated Recombination—Homologous recombination has been utilized in yeast for decades for the purposes of cloning and modifying yeast strains. More recently, the potential of the highly efficient machinery responsible for recombination has been realized towards more complex efforts, including assembling the first synthetic genome and the generation of large libraries for N-hybrid assays [41–42]. The most notable technique of this nature is transformation-associated recombination (TAR). TAR enables the capture of large genetic elements from a target organism into a selectable vector through a single transformation step. By co-transforming the target genomic DNA with a yeast compatible vector that contains “hooks” homologous to the ends of a target region, DNA sequences as large as 250 kb can be captured and isolated for heterologous gene cluster expression.

A 3-organism strategy is generally employed for TAR (Fig. 2). First, the cluster is captured by TAR in a strain of *S. cerevisiae*. Next, the captured cluster is transferred to an *E. coli* strain for vector amplification, sequencing, and any necessary modifications. Finally, the cluster is transferred to a heterologous host for gene cluster expression and compound production.

While TAR was first described in 1996, the first natural product gene cluster was captured and reported in 2010 [43–44]. In this work, the authors captured the 56 kb colibactin gene cluster from *Citrobacter koseri* using genomic DNA, as well as two additional uncharacterized PKS and NRPS clusters from eDNA cosmid libraries [43]. The eDNA approach is particularly interesting because it leveraged a core strength of TAR, the single-step capture of large genetic elements, to complement a relative weakness of cosmid libraries, the 40–50 kb insert limitation. In a more symbolic sense, this study also bridged the gap between existing techniques for natural product discovery (cosmid libraries) and new (TAR) techniques for natural product discovery. The first heterologous expression of a natural product cluster using TAR was reported later that year, with the capture and expression of a number of fluostatins from environmental DNA (eDNA) clusters in a *Streptomyces albus* host [45]. The polyketides arixanthomycin A-C were also discovered using this strategy, yielding compounds with cytotoxicity in the sub-micromolar range [46].

Bioinformatic analysis identified a hybrid fatty acid/NRPS cluster in a *Saccharomonospora* spp. that was anticipated to synthesize a halogenated daptomycin analog. Upon capture of

the cluster by TAR and expression in *S. albus*, taromycin A **(5)** was discovered and characterized. This study was the first to directly capture and express a natural product gene cluster from genomic DNA using TAR. However, initial attempts of heterologous expression using the captured cluster were unsuccessful, yielding no compound. In an effort to remove negative regulators within the cluster bounds, predicted regulatory elements were knocked out by a yeast PCR-targeting strategy (analogous to λ Red PCR-targeting), which resulted in the production of taromycin A by *S. coelicolor* M1146 [47–48].

TAR was combined with λ Red to enhance heterologous expression when capturing the bromoalterochromide cluster from *Pseudoalteromonas piscicida*. Initial production of alterochromides was 60-fold lower than the native producer. To improve this, λ Red was used to transfer the whole cluster from the initial capture vector into pETDuet-1. This placed the entire cluster downstream of the T7 promoter for inducible expression in *E. coli*, resulting in a 20-fold increase in bromoalterochromide **(6)** production relative to the endogenous promoter. According to the authors, at 34 kb, it was the largest NRPS cluster to be expressed in *E. coli* in a singular construct to date [49].

A similar strategy was used to investigate the biosynthesis of amicoumacins **(7)** from the marine isolate, *Bacillus subtilis* 1779. Following capture of the cluster by TAR, λ Red was used to knockout a peptidase gene to confirm its function which revealed the prodrug preamicoumacin when the cluster was expressed in *Bacillus subtilis* 168. Perhaps most notably, this was the first reported heterologous expression of natural products in *B. subtilis* using direct cloning, facilitated by TAR [50].

TAR capture strategies for eDNA derived clusters have also begun implementing engineering efforts as well. A cluster predicted to generate an indolotryptoline product was captured from eDNA by TAR, but no product formation was detected in the heterologous host *S. albus*. Using yeast PCR-targeting, all of the native promoters were replaced with constitutive promoters and yeast selection markers. This enabled the discovery of lazaramide, which displays cytotoxicity in the low nanomolar range [51].

As with any technology, TAR cloning has its drawbacks. While the colibactin gene cluster was amongst the first natural product gene clusters captured from genomic DNA, this attempt was unsuccessful in producing colibactin. Successful production of colibactin by a TAR strategy only came after an independent effort five years later [52]. Although TAR is able to capture clusters efficiently, some remain intractable for heterologous production. For example, the cluster for nataxazole, a cytotoxic benzoxazole, was captured by TAR but expression of the cluster caused significant growth inhibition in heterologous hosts. Moreover, some hosts could not be transformed due to the cytotoxicity of nataxazole pathway intermediates [53].

Unlike the phage-based systems, which can be performed in *E. coli* alone, TAR requires the use of *S. cerevisiae*. While this microbe is a model system for many fields, it is often a new strain for many natural products-oriented labs. As such, techniques for culturing, transforming, and genetically manipulating yeast may require significant investments for labs lacking experience with fungi or eukaryotes. Vector construction is also more complex

in TAR, as yeast replication elements and selection markers must be included in addition to the *E. coli* and heterologous host elements. The exact mechanisms of TAR are also unknown, and there may be limitations to the technology that have not been identified yet [54]. Additionally, TAR remains to be commercialized, and therefore does not yet benefit from the support that many established technologies have.

The capture of a gene cluster is often in practice not sufficient for many targets. Therefore, a number of techniques have been developed to refactor the regulatory systems controlling the cluster, successfully boosting gene expression and compound production. Another perceivable challenge is the ability of heterologous hosts to both accept and stably maintain large DNA molecules, even if very large genetic elements can be captured. However, in the current 3-organism format involving *S. cerevisiae*, *E. coli*, and the heterologous production host, TAR shows great potential for the study of natural products.

3.3.2. DNA Assembler—Like TAR, DNA assembler harnesses the homologous recombination power of yeast using short homology regions to stitch together DNA sequences (Fig. 4)[55]. While TAR focuses on capturing a natively organized gene cluster, DNA assembler refactors the entire cluster from fragments [56–57]. Each gene in the cluster is individually amplified by PCR, then joined to a promoter sequence and terminator sequence by overlap extension PCR, such that each preceding construct bears homology to the proceeding construct. These promoter-gene-terminator amplicons and linearized vector are then co-transformed into yeast and assembled into a refactored gene cluster by homologous recombination (Fig. 4)[55].

Following a proof-of-concept study in which the pathways for D-xylose utilization and zeaxanthin biosynthesis were refactored, DNA assembler was used to reconstruct the spectinabilin pathway from *Streptomyces orinoci* [55]. The design of the reconstructed cluster omitted repressor *norD* and incorporated validated strong promoters ahead of each gene for constitutive expression. As a result, spectinabilin **(8)** was produced in the heterologous host *Streptomyces lividans*. The refactoring of this known pathway represented the first application of DNA assembler towards secondary metabolites [57], which was later applied towards the characterization of three novel polycyclic tetramic acid natural products [56].

3.4. Heterologous Host Considerations

While it may seem a simple step, the selection of an appropriate host for heterologous expression is critical. The optimal expression strain is generally closely related to, and often in the same genus as, the strain the target cluster was captured from. If a closely related strain cannot be used, it is important to choose a host that is similar in codon usage and regulation of gene expression to the native host. Another important consideration is the availability of biosynthetic precursors, such as amino acids for nonribosomal peptides, and acyl-CoAs for polyketides. Finally, the host strain should be both relatively easy to culture and capable of accepting foreign DNA. Since the *Streptomyces* genus produces the lion's share of natural products, several engineered strains have been developed and are already utilized for the production of natural products. These include several strains of *S. coelicolor*

A3(2) and *S. avermitilis* [39, 58]. These strains have been engineered to remove endogenous secondary metabolite gene clusters in order to simplify the identification of heterologously produced compounds, promote carbon and nitrogen flux towards desired pathways, and to increase precursor availability [39, 58]. Additional improvements of the *S. coelicolor* A3(2) derived expression hosts were achieved by the incorporation of mutations in RNA polymerase and ribosomal protein S12, which have been shown to increase secondary metabolite production [58]. These engineered strains of *S. avermitilis* and *S. coelicolor* are suitable heterologous hosts for gene clusters captured by TAR or the phage recombination systems in addition to other strains commonly used including *S. lividans* and *S. albus*.

4. COMPARISON OF DIRECT CAPTURE WITH TRADITIONAL NATURAL PRODUCT CLONING TECHNIQUES

Traditional methods for studying natural product gene clusters have involved the creation of genomic libraries in the form of cosmids and bacterial artificial chromosomes (BACs). Forming these libraries is a well-established process that has been commercialized for use in many fields. Most laboratories are equipped to create, maintain, and implement these DNA libraries. Creating a library can still be a laborious and complex process, as it involves multiple DNA manipulation steps, phage packaging and the screening of thousands of clones, but established protocols exist for each phase. Since cosmid and BAC inserts are generated through enzymatic digest or shearing genomic DNA, there is no sequence specificity as to which fragments are incorporated into each vector, and clusters may be fragmented across multiple vectors [59]. Commercial kits are also not directly transferable to heterologous hosts. The backbones require retrofitting with replication and selection elements for maintenance in host strains, as well as transfer elements to facilitate movement of the cluster into heterologous hosts [48].

Direct capture technologies have several clear benefits when capturing gene clusters for heterologous expression. The targeted approach focuses on only the selected cluster of interest, and captures the whole cluster in a single step. In both phage-based systems and TAR, the capture vector can be pre-equipped with elements necessary for heterologous expression, which dramatically streamlines the cloning process. Construction of a cluster-specific capture vector is also relatively straightforward in both direct capture techniques. For the phage-based systems, the homology required is on the scale of 40 bp, which can be conveniently synthesized as a 5′ addition to a PCR oligonucleotide [48]. Similarly, the homology required for the TAR system requires less than 1 kb (documented down to less than 60 bp), which eliminates the need for long PCRs to generate homology arms (Fig. 2) [60].

Phage-based recombination is well studied and widely implemented, but its use is often limited to capture gene clusters less than 40 kb in size. TAR, by comparison, is able to overcome these size limitations, but requires a much more specialized approach. Unlike the widely utilized *E. coli, S. cerevisiae* is less routinely used as a tool for many natural products laboratories. Protocols for TAR are still being developed, and the methodologies in design and implementation vary between labs.

In many ways, library methods and directed capture are complementary approaches. A cosmid/BAC library can be generated to capture an entire genome or metagenome, and direct capture can be then used to assemble the desired cluster in a single vector, as demonstrated by the capture of numerous clusters from eDNA libraries, as noted earlier [43, 45–46, 51, 61].

Ultimately, direct capture takes advantage of the growing wealth of sequencing information, enabling a true genomics-guided approach. However, cosmid/BAC libraries remain useful technologies when sequencing information is unavailable or low-resolution.

5. SUMMARY

In this review we have described how these homologous recombination systems have been applied as a form of direct capture for the study of secondary metabolites since 2010. These efforts not only led to heterologous production of compounds of interest, but also enabled the discovery of novel compounds.

As of this writing, more successful TAR studies have been reported than the combined λ Red/ET recombination efforts. Although direct capture has simplified gene cluster isolation, further manipulations of the cluster are still beneficial and sometime necessary for compound production. Repressor deletions and promoter exchange are just two possible modifications that are afforded by homologous recombination. Both the *E. coli* based system and the *S. cerevisiae* system have seen success in natural product gene cluster capture; however, TAR has been used more successfully for cluster capture while phage recombination is more readily used for cluster modification following capture.

6. OUTLOOK – RECENT INNOVATIONS AND PIONEERING THE FUTURE OF NATURAL PRODUCTS

Homologous recombination based techniques hold promise for the future of natural product discovery and for unlocking cryptic gene clusters. The applications of TAR in the capture of gene clusters from environmental DNA have allowed for analysis of clusters that are natively harbored by ‘unculturable’ bacteria. Future applications of TAR might directly capture gene clusters from metagenomic DNA, without the need of building cosmid libraries first. In either case, TAR has great potential as a technology to discover novel compounds from environmental bacteria whose genes have been previously deemed inaccessible.

Besides phage recombination in *E. coli* and TAR in yeast, *Bacillus subtilis* is another model platform that may serve as a suitable vector for the capture of natural product gene clusters. The 4.2 Mb *B. subtilis* 168 genome is known as the *B. subtilis* genome-based manipulation (BGM) vector [62]. The BGM vector has been used in the reconstruction of an entire 3.5 Mb genome from many individual and overlapping pieces, in a fashion called the domino method or “domino cloning” [62–63]. Furthermore, *B. subtilis* has a high homologous recombination rate and natural competence, making it an ideal host for direct capture [63–65]. *B. subtilis* has been used as an effective host for the heterologous expression of natural product gene clusters, including the expression of an amicoumacin cluster captured directly

by TAR as well as the nisin, rhizocticins, and polymyxin clusters captured indirectly [50, 66–68]. Despite these advantages, *B. subtilis* has yet to be exploited for direct capture applications to discover new natural products from cryptic gene clusters. *B. subtilis* has the potential to become a powerful tool for the capture of natural product gene clusters in a similar fashion to phage recombination and TAR, especially for AT-rich bacteria [61–62].

We anticipate that emerging genomic editing tools will further the discovery and heterologous production of natural products. Recently, two genome modification techniques have been developed for the *Streptomyces* genus, including the SceI meganuclease of *S. cerevisiae* and the CRISPR system, which function by creating lethal double stranded breaks [69–72]. A homology directed repair cassette facilitates the repair of these breaks and avoids lethality. Deletions and insertions of individual genes are well-documented but using either system to insert an entire cluster remains to be examined. In principle, smaller gene clusters would be better tolerated. The homology directed repair cassettes containing the gene clusters of interest could be sourced from a TAR or phage-based capture system. The combination of techniques will likely provide opportunities for multiplexing, dramatically accelerating the rate of genomics-guided discovery of natural products.

Although all of these techniques aim to make cryptic gene clusters more readily accessible there is no singular recipe for success; each cluster and system can be considered ‘unique,’ and it is likely that some cryptic clusters await further technological developments for their unlocking. The unprecedented ability to capture large genetic elements in a targeted approach, combined with recent methods to modify these clusters, has the potential to change the way natural product gene clusters are studied. These techniques may open up new branches in the study of microbial natural products and allow researchers to investigate clusters previously thought inaccessible, either due to degree of complexity or from poor understanding of the native organism.

Acknowledgments

Research in the authors’ laboratory is supported by the University of North Carolina at Chapel Hill and NIH grant 5R00GM099904-04 (B.L.).

References

1. Medema MH, Trefzer A, Kovalchuk A, van den Berg M, Müller U, Heijne W, Wu L, Alam MT, Ronning CM, Nierman WC, Bovenberg RAL, Breitling R, Takano E. The sequence of a 1.8-Mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. Genome Biol Evol. 2010; 2:212–224. [PubMed: 20624727]
2. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. antiSMASH, rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011; 39(Web Server issue):W339–46. [PubMed: 21672958]
3. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos, an information aesthetic for comparative genomics. Genome Res. 2009; 19(9):1639–45. [PubMed: 19541911]
4. Luo Y, Cobb RE, Zhao H. Recent advances in natural product discovery. Curr Opin Biotechnol. 2014; 30:230–237. [PubMed: 25260043]

5. Unsin, C-M.; Rajski, S.; Shen, B. The role of genetic engineering in natural product-based anticancer drug discovery. In: Koehn, FE., editor. Natural Products and Cancer Drug Discovery. Springer; New York: 2013. p. 175-191.
6. Zerikly M, Challis GL. Strategies for the discovery of new natural products by genome mining. ChemBioChem. 2009; 10:625–633. [PubMed: 19165837]
7. Ochi K, Hosaka T. New strategies for drug discovery, activation of silent or weakly expressed microbial gene clusters. Appl Microbiol Biotechnol. 2013; 97:87–98. [PubMed: 23143535]
8. Seyedsayamdost, MR.; Traxler, MF.; Clardy, J.; Kolter, R. Methods Enzymol. Vol. 517. Elsevier; 2012. Old meets new, Using interspecies interactions to detect secondary metabolite production in Actinomycetes; p. 89-109.
9. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants, the Keio collection. Mol Syst Biol. 2006; 2:0008. [PubMed: 16738554]
10. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Guldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kotter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD, Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang CY, Ward TR, Wilhelmy J, Winzeler EA, Yang Y, Yen G, Youngman E, Yu K, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW, Johnston M. Functional profiling of the *Saccharomyces cerevisiae* genome. Nature. 2002; 418(6896):387–91. [PubMed: 12140549]
11. Walsh CT, Fischbach MA. Natural products version 2.0, connecting genes to molecules. J Am Chem Soc. 2010; 132(8):2469–93. [PubMed: 20121095]
12. Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, Birren BW, Takano E, Sali A, Linington RG, Fischbach MA. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell. 2014; 158(2):412–21. [PubMed: 25036635]
13. Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND. SMURF, Genomic mapping of fungal secondary metabolite clusters. Fungal Genet Biol. 2010; 47(9):736–41. [PubMed: 20554054]
14. Stachelhaus T, Marahiel MA. Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. J Biol Chem. 1995; 270(11):6163–9. [PubMed: 7534306]
15. Stachelhaus T, Mootz HD, Marahiel MA. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem Biol. 1999; 6(8):493–505. [PubMed: 10421756]
16. Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L. Modular organization of genes required for complex polyketide biosynthesis. Science. 1991; 252(5006):675–9. [PubMed: 2024119]
17. Yadav G, Gokhale RS, Mohanty D. Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. J Mol Biol. 2003; 328(2):335–63. [PubMed: 12691745]
18. Gross H, Stockwell VO, Henkels MD, Nowak-Thompson B, Loper JE, Gerwick WH. The genomisotopic approach, a systematic method to isolate products of orphan biosynthetic gene clusters. Chem Biol. 2007; 14(1):53–63. [PubMed: 17254952]
19. Davidsen JM, Bartley DM, Townsend CA. Non-ribosomal propeptide precursor in nocardicin A biosynthesis predicted from adenylation domain specificity dependent on the MbtH family protein NocI. J Am Chem Soc. 2013; 135(5):1749–59. [PubMed: 23330869]
20. Gaudelli NM, Long DH, Townsend CA. beta-Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. Nature. 2015; 520(7547):383–7. [PubMed: 25624104]
21. Inglis DO, Binkley J, Skrzypek MS, Arnaud MB, Cerqueira GC, Shah P, Wymore F, Wortman JR, Sherlock G. Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of *Aspergillus nidulans*, *A. fumigatus*, *A. niger* and *A. oryzae*. BMC Microbiol. 2013; 13:91. [PubMed: 23617571]

22. Revill WP, Bibb MJ, Hopwood DA. Purification of a malonyltransferase from *Streptomyces coelicolor* A3(2) and analysis of its genetic determinant. J Bacteriol. 1995; 177:3946–3952. [PubMed: 7608065]
23. Wesener SR, Potharla VY, Cheng YQ. Reconstitution of the FK228 biosynthetic pathway reveals cross talk between modular polyketide synthases and fatty acid synthase. Appl Environ Microbiol. 2011; 77:1501–1507. [PubMed: 21183648]
24. Wang C, Henkes LM, Doughty LB, He M, Wang D, Meyer-Almes FJ, Cheng YQ. Thailandepsins, Bacterial products with potent histone deacetylase inhibitory activities and broad-spectrum antiproliferative activities. J Nat Prod. 2011; 74:2031–2038. [PubMed: 21793558]
25. Buell CR, Joardar V, Lindeberg M, Selengut J, Paulsen IT, Gwinn ML, Dodson RJ, Deboy RT, Durkin AS, Kolonay JF, Madupu R, Daugherty S, Brinkac L, Beanan MJ, Haft DH, Nelson WC, Davidsen T, Zafar N, Zhou L, Liu J, Yuan Q, Khouri H, Fedorova N, Tran B, Russell D, Berry K, Utterback T, Van Aken SE, Feldblyum TV, D'Ascenzo M, Deng WL, Ramos AR, Alfano JR, Cartinhour S, Chatterjee AK, Delaney TP, Lazarowitz SG, Martin GB, Schneider DJ, Tang X, Bender CL, White O, Fraser CM, Collmer A. The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000. Proc Natl Acad Sci US A. 2003; 100:10181–10186.
26. Vingadassalon A, Lorieux F, Juguet M, Le Goff G, Gerbaud C, Pernodet JL, Lautru S. Natural combinatorial biosynthesis involving two clusters for the synthesis of three pyrrolamides in *Streptomyces netropsis*. ACS Chem Biol. 2015; 10:601–610. [PubMed: 25415678]
27. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC, Cruz-Morales P, Duddela S, Düsterhus S, Edwards DJ, Fewer DP, Garg N, Geiger C, Gomez-Escribano JP, Greule A, Hadjithomas M, Haines AS, Helfrich EJN, Hillwig ML, Ishida K, Jones AC, Jones CS, Jungmann K, Kegler C, Kim HU, Kötter P, Krug D, Masschelein J, Melnik AV, Mantovani SM, Monroe EA, Moore M, Moss N, Nützmann HW, Pan G, Pati A, Petras D, Reen FJ, Rosconi F, Rui Z, Tian Z, Tobias NJ, Tsunematsu Y, Wiemann P, Wyckoff E, Yan X, Yim G, Yu F, Xie Y, Aigle B, Apel AK, Balibar CJ, Balskus EP, Barona-Gómez F, Bechthold A, Bode HB, Borriss R, Brady SF, Brakhage AA, Caffrey P, Cheng YQ, Clardy J, Cox RJ, De Mot R, Donadio S, Donia MS, van der Donk WA, Dorrestein PC, Doyle S, Driessen AJM, Ehling-Schulz M, Entian KD, Fischbach MA, Gerwick L, Gerwick WH, Gross H, Gust B, Hertweck C, Höfte M, Jensen SE, Ju J, Katz L, Kaysser L, Klassen JL, Keller NP, Kormanec J, Kuipers OP, Kuzuyama T, Kyrpides NC, Kwon HJ, Lautru S, Lavigne R, Lee CY, Linquan B, Liu X, Liu W, Luzhetskyy A, Mahmud T, Mast Y, Méndez C, Metsä-Ketelä M, Micklefield J, Mitchell DA, Moore BS, Moreira LM, Müller R, Neilan BA, Nett M, Nielsen J, O'Gara F, Oikawa H, Osbourn A, Osburne MS, Ostash B, Payne SM, Pernodet JL, Petricek M, Piel J, Ploux O, Raaijmakers JM, Salas JA, Schmitt EK, Scott B, Seipke RF, Shen B, Sherman DH, Sivonen K, Smanski MJ, Sosio M, Stegmann E, Süssmuth RD, Tahlan K, Thomas CM, Tang Y, Truman AW, Viaud M, Walton JD, Walsh CT, Weber T, van Wezel GP, Wilkinson B, Willey JM, Wohlleben W, Wright GD, Ziemert N, Zhang C, Zotchev SB, Breitling R, Takano E, Glöckner FO. Minimum information about a biosynthetic gene cluster. Nature Chem Biol. 2015; 11:625–631. [PubMed: 26284661]
28. Murphy KC. Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*. J Bacteriol. 1998; 180:2063–2071. [PubMed: 9555887]
29. Zhang Y, Buchholz F, Muyrers JP, Stewart AF. A new logic for DNA engineering using recombination in *Escherichia coli*. Nat Genet. 1998; 20(2):123–8. [PubMed: 9771703]
30. Zhang Y, Muyrers JP, Testa G, Stewart AF. DNA cloning by homologous recombination in *Escherichia coli*. Nat Biotechnol. 2000; 18(12):1314–7. [PubMed: 11101815]
31. Muyrers JPP, Zhang YM, Stewart AF. Techniques, Recombinogenic engineering - new options for cloning and manipulating DNA. Trends Biochem Sci. 2001; 26(5):325–331. [PubMed: 11343926]
32. Pines G, Freed EF, Winkler JD, Gill RT. Bacterial recombineering, Genome engineering via phage-based homologous recombination. ACS Synth Biol. 2015; 4(11):1176–1185. [PubMed: 25856528]
33. Murphy KC. Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of *Escherichia coli* RecBCD enzyme. J Bacteriol. 1991; 173(18):5808–21. [PubMed: 1653221]
34. Fu J, Bian X, Hu S, Wang H, Huang F, Seibert PM, Plaza A, Xia L, Muller R, Stewart AF, Zhang Y. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. Nat Biotechnol. 2012; 30(5):440–6. [PubMed: 22544021]

35. Maresca M, Erler A, Fu J, Friedrich A, Zhang Y, Stewart AF. Single-stranded heteroduplex intermediates in lambda Red homologous recombination. BMC Mol Biol. 2010; 11:54. [PubMed: 20670401]
36. Lim SI, Min BE, Jung GY. Lagging strand-biased initiation of red recombination by linear double-stranded DNAs. J Mol Biol. 2008; 384(5):1098–105. [PubMed: 18983848]
37. Bian X, Huang F, Stewart FA, Xia L, Zhang Y, Muller R. Direct cloning, genetic engineering, and heterologous expression of the syringolin biosynthetic gene cluster in *E.coli* through Red/ET recombineering. Chembiochem. 2012; 13(13):1946–52. [PubMed: 22851214]
38. Komatsu M, Komatsu K, Koiwai H, Yamada Y, Kozone I, Izumikawa M, Hashimoto J, Takagi M, Omura S, Shinya K, Cane DE, Ikeda H. Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. ACS Synth Biol. 2013; 2(7): 384–96. [PubMed: 23654282]
39. Komatsu M, Uchiyama T, Omura S, Cane DE, Ikeda H. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. Proc Natl Acad Sci USA. 2010; 107(6): 2646–51. [PubMed: 20133795]
40. Yin J, Zhu H, Xia L, Ding X, Hoffmann T, Hoffmann M, Bian X, Muller R, Fu J, Stewart AF, Zhang Y. A new recombineering system for *Photorhabdus* and *Xenorhabdus*. Nucleic Acids Res. 2015; 43(6):e36. [PubMed: 25539914]
41. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA 3rd, Smith HO, Venter JC. Creation of a bacterial cell controlled by a chemically synthesized genome. Science. 2010; 329(5987):52–6. [PubMed: 20488990]
42. Romanini DW, Peralta-Yahya P, Mondol V, Cornish VW. A heritable recombination system for synthetic Darwinian evolution in yeast. ACS Synth Biol. 2012; 1(12):602–9. [PubMed: 23412545]
43. Kim JH, Feng Z, Bauer JD, Kallifidas D, Calle PY, Brady SF. Cloning large natural product gene clusters from the environment, Piecing environmental DNA gene clusters back together with TAR. Biopolymers. 2010; 93:833–844. [PubMed: 20577994]
44. Larionov V, Kouprina N, Graves J, Chen XN, Korenberg JR, Resnick MA. Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. Proc Natl Acad Sci USA. 1996; 93(1):491–496. [PubMed: 8552668]
45. Feng Z, Kim JH, Brady SF. Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived type II PKS gene cluster. J Am Chem Soc. 2010; 132(34):11902–3. [PubMed: 20690632]
46. Kang HS, Brady SF. Arixanthomycins A-C, Phylogeny-guided discovery of biologically active eDNA-derived pentangular polyphenols. ACS Chem Biol. 2014; 9(6):1267–72. [PubMed: 24730509]
47. Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC, Moore BS. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc Natl Acad Sci USA. 2014; 111(5):1957–62. [PubMed: 24449899]
48. Gust, B.; Kieser, T.; Chater, KF. Streptomyces coelicolor. The John Innes Centre; Norwich, United Kingdom: 2002. REDIRECT technology, PCR-targeting system.
49. Ross AC, Gulland LE, Dorrestein PC, Moore BS. Targeted capture and heterologous expression of the *Pseudoalteromonas* alterochromide gene cluster in *Escherichia coli* represents a promising natural product exploratory platform. ACS Synth Biol. 2015; 4(4):414–20. [PubMed: 25140825]
50. Li Y, Li Z, Yamanaka K, Xu Y, Zhang W, Vlamakis H, Kolter R, Moore BS, Qian PY. Directed natural product biosynthesis gene cluster capture and expression in the model bacterium *Bacillus subtilis*. Sci Rep. 2015; 5:9383. [PubMed: 25807046]
51. Montiel D, Kang HS, Chang FY, Charlop-Powers Z, Brady SF. Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. Proc Natl Acad Sci USA. 2015; 112(29):8953–8. [PubMed: 26150486]
52. Li ZR, Li Y, Lai JY, Tang J, Wang B, Lu L, Zhu G, Wu X, Xu Y, Qian PY. Critical intermediates reveal new biosynthetic events in the enigmatic colibactin pathway. Chembiochem. 2015; 16(12): 1715–9. [PubMed: 26052818]

53. Cano-Prieto C, Garcia-Salcedo R, Sanchez-Hidalgo M, Brana AF, Fiedler HP, Mendez C, Salas JA, Olano C. Genome mining of *Streptomyces* sp. Tu 6176; Characterization of the nataxazole biosynthesis pathway. Chembiochem. 2015; 16(10):1461–73. [PubMed: 25892546]
54. Kawai S, Hashimoto W, Murata K. Transformation of *Saccharomyces cerevisiae* and other fungi. Bioengineered Bugs. 2010; 1:395–403. [PubMed: 21468206]
55. Shao Z, Zhao H, Zhao H. DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. Nucleic Acids Res. 2009; 37(2):e16. [PubMed: 19074487]
56. Luo Y, Huang H, Liang J, Wang M, Lu L, Shao Z, Cobb RE, Zhao H. Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. Nat Commun. 2013; 4:2894. [PubMed: 24305602]
57. Shao Z, Rao G, Li C, Abil Z, Luo Y, Zhao H. Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. ACS Synth Biol. 2013; 2(11):662–9. [PubMed: 23968564]
58. Gomez-Escribano JP, Bibb MJ. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. Microb Biotechnol. 2011; 4(2):207–15. [PubMed: 21342466]
59. Sambrook, J. Molecular Cloning, A Laboratory Manual. 3. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, N.Y: 2001. p. 999
60. Noskov VN, Koriabine M, Solomon G, Randolph M, Barrett JC, Leem SH, Stubbs L, Kouprina N, Larionov V. Defining the minimal length of sequence homology required for selective gene isolation by TAR cloning. Nucleic Acids Res. 2001; 29(6):E32. [PubMed: 11239009]
61. Kallifidas, D.; Brady, SF. Chapter Eleven - Reassembly of Functionally Intact Environmental DNA-Derived Biosynthetic Gene Clusters. In: Hopwood, DA., editor. Methods Enzymol. Vol. 517. Academic Press; 2012. p. 225-239.
62. Itaya M, Fujita K, Kuroki A, Tsuge K. Bottom-up genome assembly using the *Bacillus subtilis* genome vector. Nat Methods. 2008; 5(1):41–3. [PubMed: 18066072]
63. Itaya M, Tsuge K, Koizumi M, Fujita K. Combining two genomes in one cell, stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. Proc Natl Acad Sci USA. 2005; 102(44):15971–6. [PubMed: 16236728]
64. Itaya M, Nagata T, Shiroishi T, Fujita K, Tsuge K. Efficient cloning and engineering of giant DNAs in a novel *Bacillus subtilis* genome vector. J Biochem. 2000; 128(5):869–75. [PubMed: 11056400]
65. Dubnau D. Genetic competence in *Bacillus subtilis*. Microbiol Rev. 1991; 55:395–424. [PubMed: 1943994]
66. Yuksel S, Hansen JN. Transfer of nisin gene cluster from *Lactococcus lactis* ATCC 11454 into the chromosome of *Bacillus subtilis* 168. Appl Microbiol Biotechnol. 2007; 74(3):640–9. [PubMed: 17143619]
67. Borisova SA, Circello BT, Zhang JK, van der Donk WA, Metcalf WW. Biosynthesis of rhizocticins, antifungal phosphonate oligopeptides produced by *Bacillus subtilis* ATCC6633. Chem Biol. 2010; 17(1):28–37. [PubMed: 20142038]
68. Choi SK, Park SY, Kim R, Kim SB, Lee CH, Kim JF, Park SH. Identification of a polymyxin synthetase gene cluster of *Paenibacillus polymyxa* and heterologous expression of the gene in *Bacillus subtilis*. J Bacteriol. 2009; 191(10):3350–8. [PubMed: 19304848]
69. Cobb RE, Wang Y, Zhao H. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. ACS Synth Biol. 2015; 4(6):723–8. [PubMed: 25458909]
70. Siegl T, Petzke L, Welle E, Luzhetskyy A. I-SceI endonuclease, a new tool for DNA repair studies and genetic manipulations in *Streptomyces*. Appl Microbiol Biotechnol. 2010; 87(4):1525–32. [PubMed: 20473607]
71. Siegl T, Luzhetskyy A. Actinomycetes genome engineering approaches. Antonie van Leeuwenhoek. 2012; 102:503–516. [PubMed: 22926833]
72. Fernandez-Martinez LT, Bibb MJ. Use of the meganuclease I-SceI of *Saccharomyces cerevisiae* to select for gene deletions in actinomycetes. Sci Rep. 2014; 4:7100. [PubMed: 25403842]

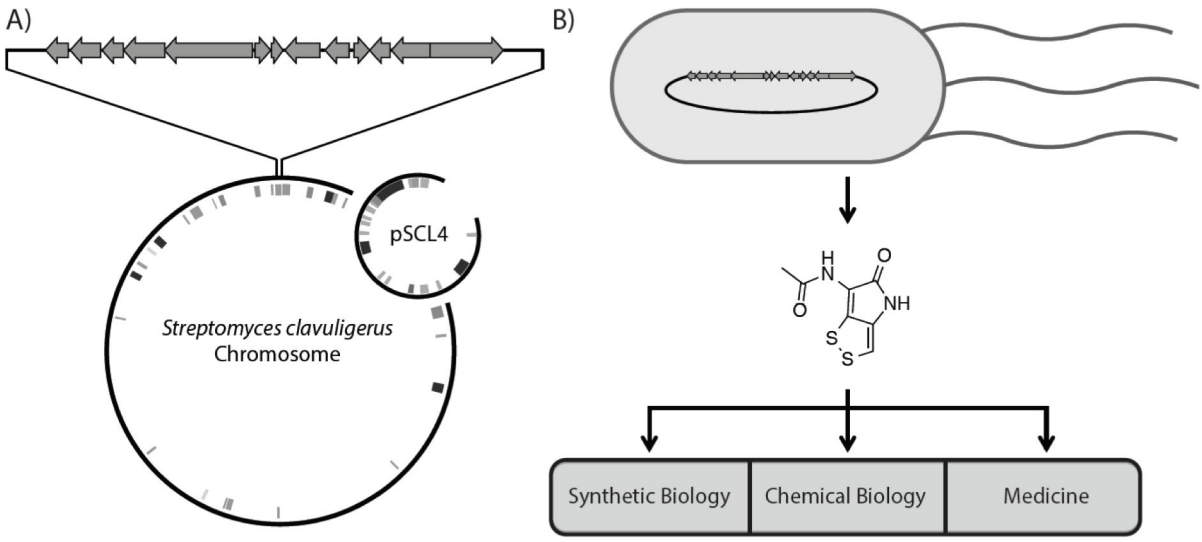

Fig. (1).

Applications of genome mining and heterologous expression towards natural products discovery. **A)** Bioinformatic analysis of microbial genomes can identify natural product gene clusters, such as that of holomycin. Gene clusters were identified by antiSMASH analysis [2], and visualized with Circos [3]. **B)** Heterologous expression of these cultures facilitates the elucidation of their products, which can be used towards a variety of scientific and medicinal efforts.

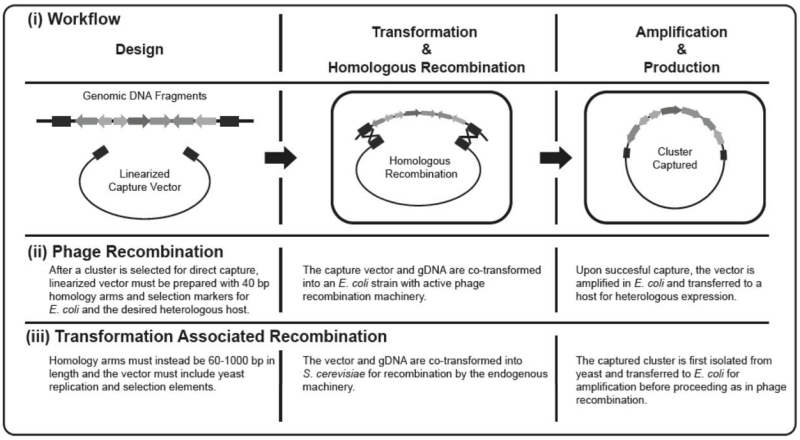

Fig. (2).

Comparisons of steps in phage recombination and transformation-associated recombination for direct capture. (i) The generalized steps in direct capture and heterologous natural product gene cluster expression. (ii) The specific steps in direct capture by phage recombination. (iii) Differences of transformation-associated recombination from phage recombination.

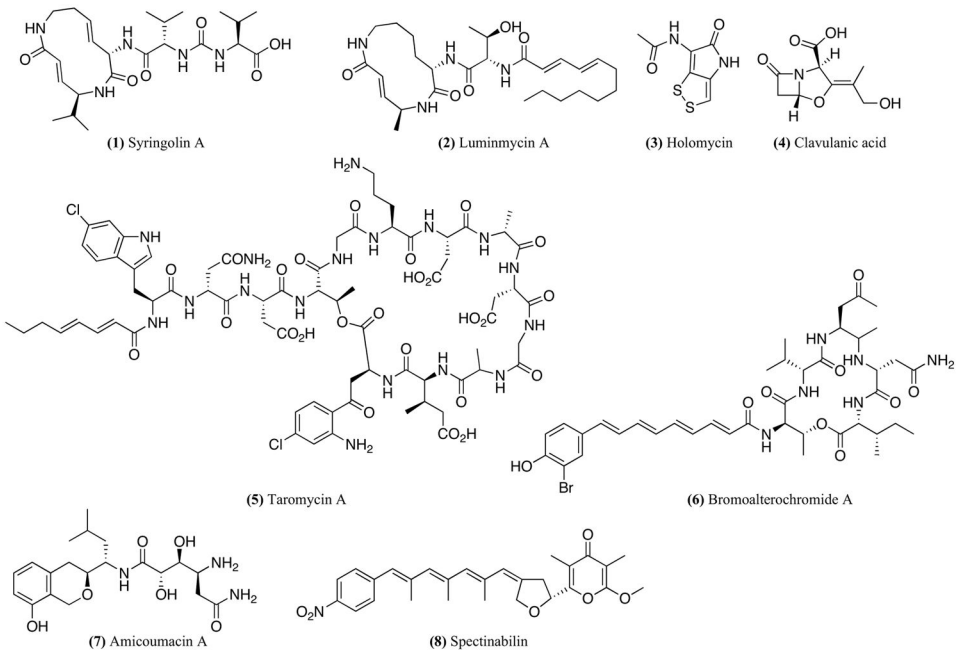

Fig. (3).

Selected structures of compounds directly captured and heterologously produced. **(1)** syringolin A, **(2)** luminmycin A, **(3)** holomycin, **(4)** clavulanic acid, **(5)** taromycin A, **(6)** bromoalterochromide A, **(7)** amicoumacin A, **(8)** spectinabilin.

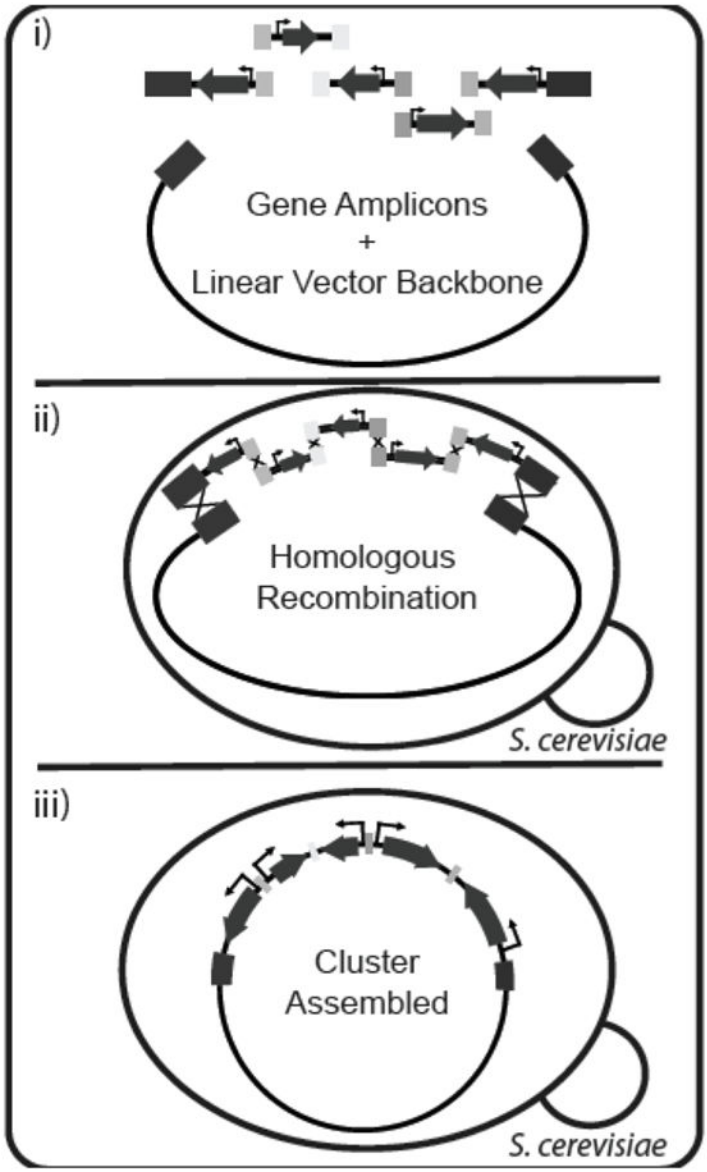

Fig. (4).

DNA assembler scheme. (i) DNA fragments are amplified to bear homology to each proceeding and preceding molecule (color blocks). The vector is also linearized and designed with homology to both ends of the cluster. (ii) *S. cerevisiae* is transformed with the DNA fragments whereupon homologous recombination assembles the cluster as in (iii).