Published in final edited form as:
Curr Protoc Chem Biol. ; 6(2): 65–100. doi:10.1002/9780470559277.ch130191.

Manipulating Natural Product Biosynthetic Pathways via DNA Assembler

Zengyi Shao[1,+] and **Huimin Zhao**[1,2,3,*]

[1]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801

[2]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801

[3]Departments of Chemistry, Biochemistry, and Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801

Abstract

DNA assembler is an efficient synthetic biology method for constructing and manipulating biochemical pathways. The rapidly increasing number of sequenced genomes provides a rich source for discovery of gene clusters involved in synthesizing new natural products. However, both discovery and economical production are hampered by our limited knowledge in manipulating most organisms and the corresponding pathways. By taking advantage of yeast *in vivo* homologous recombination, DNA assembler synthesizes an entire expression vector containing the target biosynthetic pathway and the genetic elements needed for DNA maintenance and replication. Here we use the spectinabilin clusters originated from two hosts as examples to illustrate the guidelines of using DNA assembler for cluster characterization and silent cluster activation. Such strategies offer unprecedented versatility in cluster manipulation, bypass the traditional laborious strategies to elicit pathway expression, and provide a new platform for *de novo* cluster assembly and genome mining for discovering new natural products.

Keywords

natural product; gene cluster; regulation of natural product biosynthesis; pathway engineering; synthetic biology; cryptic pathway; DNA assembler

INTRODUCTION

This article describes how to take full advantage of DNA assembler to study natural product biosynthesis. The majority of existing antibacterial and anticancer drugs are natural products or their derivatives (Li and Vederas, 2009; Newman and Cragg, 2012). Over the last two decades, the complete genome sequences of more than 2,000 organisms have been determined, and more than 10,000 organisms being sequenced. Genome mining has revealed that the natural products that have been characterized are merely “the tip of the iceberg”,

[*]To whom correspondence should be addressed. Phone: (217) 333-2631. Fax: (217) 333-5052. zhao5@illinois.edu.
[+]Current address: Department of Chemical and Biological Engineering, Iowa State University, Ames, IA, 50011

whilst plenty of metabolites await discovery (Abed et al., 2009; Bode and Muller, 2006; Brakhage and Schroeckh, 2011; Ohnishi et al., 2008; Omura et al., 2001). However, the characterization and discovery of natural products are often hampered by our limited ability to manipulate the corresponding biosynthetic pathways. Most native producers cannot be cultivated in the laboratory and lack tools for genetic manipulation (Davies, 1999). Utilizing well-characterized organisms as heterologous hosts alleviates these problems to a certain extent, but still suffers from problems with promoter recognition and gene expression, and the difficulty in performing sophisticated genetic modifications. The biosynthesis of natural products is highly regulated and gene clusters often remain silent until suitable conditions are met (Bibb and Hesketh, 2009; Bibb, 2005; Brakhage et al., 2008; van Wezel et al., 2009). Our current understanding of regulation hierarchy is not complete enough to predict the functions of all the regulatory elements. Thus, what are highly desirable in natural product studies are efficient approaches to rapidly manipulate large gene clusters and generic strategies that simplify regulation complexity and avoid the need to tailor efforts according to specific clusters. Here we use the spectinabilin gene clusters from *Streptomyces spectabilis* and *Streptomyces orinoci* as two examples to illustrate (1) how to apply DNA assembler to perform scar-less gene disruption and site-directed mutagenesis for rapid gene cluster characterization and (2) how to reconstruct an artificial natural product gene clusterin order to decouple its expression from native sophisticated regulation cascades.

Spectinabilin is a nitro-phenyl containing polyketide that exhibits antimalarial and antiviral activities (Isaka et al., 2002; Kakinuma et al., 1976). The two clusters from *S. spectabilis* and *S. orinoci* both consist of four polyketide synthases (PKSs), and the catalytic proteins from these PKSs share very high sequence homologies with each other (Figure 1). Examining the two clusters revealed two major differences. First, the *S. spectabilis* cluster (called as spn cluster or spn pathway hereafter) contains three extra genes, *spnK*, *spnL*, and *spnM*, which are missing from the *S. orinoci* cluster (referred to as the “nor cluster” or “nor pathway” hereafter) (Choi et al., 2010). Second, these two clusters undergo different regulations as the spn cluster contains *spnD*, which encodes an activator whereas the nor cluster contains *norD*, which encodes a repressor. When directly cloned into *Streptomyces lividans*, only the spn cluster could produce spectinabilin (Choi et al., 2010; Traitcheva, 2006). To study the spectinabilin biosynthetic mechanism and identify the reason for the nor cluster being silent in *S. lividans*, in the following two protocols, we first use DNA assembler to investigate the essentialness of *spnK*, *spnL*, and *spnM* in spectinabilin biosynthesis, then refactor the nor pathway by replacing the native promoters with heterologous promoters and thus activate the silent cluster.

STRATEGIC PLANNING

The strategies are designed based on heterologous host expression, and two circumstances exist (Figure 2). In the first case, when the promoters from a native producer are recognized in a heterologous host, the direct cloning strategy will be used and an expression plasmid consisting of a pathway module and three helper modules will be assembled. The target gene cluster will be split into several pathway fragments, composing the pathway module. Each pathway fragment has a length of up to 6 kb. Pathway fragments are amplified by polymerase chain reaction (PCR) from the isolated genomic DNA of the native producer if it

is cultivable or chemically synthesized *de novo*. Three helper modules carrying the genetic elements needed for DNA maintenance and replication in *S. cerevisiae* (assembly host), *E. coli* (DNA enrichment host), and the target heterologous expression host, are amplified from the corresponding vectors. Here the minimal genetic elements include a selection marker used in each host, and a functional element denoting whether the plasmid will be maintained in cytosol or integrated to chromosome. Since PCR primers are designed to generate overlap regions between adjacent fragments, co-transformation of these fragments into *S. cerevisiae* will allow them to be assembled into a single DNA molecule through homologous recombination. The isolated plasmids are subsequently transformed to *E. coli* for plasmid enrichment and verification, and the correct construct is transformed into the desired host for heterologous expression of the target pathway.

However, in some cases, direct transferring of a cluster from a native producer to a heterologous host results in unrecognized promoters and thus unsuccessful expression. To overcome this problem, the gene cluster needs to be reconstructed. As shown in Figure 2, the entire pathway module in the first case should be split into a few promoter modules and gene modules whereas the helper modules can be maintained. The promoter modules are set up based on the choice of the expression host. Promoters can be chosen from the well-studied ones reported in literature (which unfortunately have not been reported in literature for most organisms) or cloned from the potential promoter regions of endogenous housekeeping genes in the target expression host. Other organisms closely related to the target expression host can also be considered as effective sources for identifying promoters that can be recognized. After the promoter modules and the helper modules are set up, individual genes can be amplified or synthesized and plugged into each gene module. Overlaps are designed between adjacent fragments so that the reconstructed pathway can be assembled using the same procedures as in the direct cloning strategy. The expression of the pathway genes can be confirmed *via* real-time PCR, SDS-PAGE and Western blot (if a gene is tagged). The product can be identified using analytical methods such as LC-MS or biological activity assay. In some cases, inducible expression is desired because the final product or some of the biosynthetic intermediates are toxic to the expression host. An inducible promoter can be inserted upstream of the gene encoding the enzyme catalyzing the first step in the biosynthesis. From another point of view, in nature, secondary metabolites are mostly synthesized in stationary growth phase, when native producers do not need a lot of resources to support their primary metabolism. Therefore, the competition between primary metabolite biosynthesis and the target secondary metabolite biosynthesis for the precursors is at least partially relieved by the inclusion of an inducible promoter. Using such a modular design, the sophisticated regulation embedded in individual clusters is removed and replaced by a set of regulation that is predictable, easy to manipulate, and not specifically linked to any gene cluster. As a side note, individual genes can be chemically synthesized *de novo* with optimized codons, resulting in a translationally improved pathway. All of these genetic manipulations are extremely difficult to accomplish by conventional methods.

In both methods mentioned above, in order to ensure high assembly efficiency, it is critical to design appropriate lengths of overlaps between adjacent fragments. For larger gene

clusters (more than 20 kb), increasing the length of overlap is necessary for obtaining a decent efficiency. In the direct cloning strategy, as shown in Figure 3a, overlaps of approximately 400 bp between internal adjacent pathway fragments can be generated easily. For example, the forward primer for amplifying the second pathway fragment can be located ~400 bp upstream of the annealing position of the reverse primer for amplifying the first pathway fragment. In the cluster reconstructing strategy, promoter modules and gene modules are separately prepared, and their in-between overlaps are generated by adding tails to primers, thus they cannot be very long, and usually have lengths of 80 bp (Figure 3b). The overlaps between helper modules and adjacent fragments are 40 bp or 80 bp. We encountered difficulties in amplifying the *S. cerevisiae* helper fragment, and in the end, the correct product was only obtained by using the pair of primers without any tails. As a result, the *S. cerevisaie* helper fragment only overlaps with the last gene module and the adjacent helper fragment with overlaps of 40 bp. Overlaps of 80 bp are generated between the other helper fragments and their neighbors. In both protocols, *S. lividans*, a laboratory strain that has been used extensively for studying gene clusters from Streptomycetes is chosen as a target heterologous host owing to its high conjugative DNA transfer efficiency and the availability of several genetic manipulation tools (Baltz, 2010; Felnagle et al., 2007; Shao et al., 2011; Woodyer et al., 2006).

BASIC PROTOCOL 1. CHARACTERIZATION OF THE SPN GENE CLUSTER VIA DIRECT CLONING

As mentioned above, the spn gene cluster from *S. spectabilis* was previously successfully expressed in *S. lividans* (Choi et al., 2010), indicating the native promoters can be recognized by the heterologous host. When applying the direct cloning strategy, we start by assembling a wild type pathway. In our initial trials, the ~45 kb spn pathway was split into 11 pathway fragments and each had a length of 4–6 kb. In principle, the number of pathway fragments to be assembled should be minimized but low yield or even failure could occur more frequently when the PCR product is relatively long (> 6 kb). However, a low assembly efficiency (<10%) was observed and most of the constructs underwent severe gene deletion(s). FThe failure was due to the high sequence similarity among PKS domains. To address this issue, a modified two-step assembly strategy was devised (Figure 4), through which we were able to assemble this biosynthetic pathway of 45 kb with an efficiency of 30%.

Materials

Wizard® Genomic DNA Purification Kit (Promega, Madison, WI)

FailSafe PCR 2×PreMix G (EPICENTRE Biotechnologies, Madison, WI, USA)

Phusion DNA polymerase and GC Reaction buffer (New England Biolabs, Beverly, MA, USA)

Dimethyl sulfoxide (DMSO)

dNTP premix containing 10 mM each nucleotide

QIAquick Gel Extraction Kit (QIAGEN, Valencia, CA, USA)

Nanodrop (ThermoFisher Scientific, Waltham, MA, USA)

0.2 cm-gap electroporation cuvette (ThermoFisher Scientific, Waltham, MA, USA)

3 M sodium acetate

10 mg/mL glycogen

Ethanol

YPAD

1 M sorbitol

SC-Ura plates

SC-Ura liquid medium

Zymoprep II kit (Zymo Research, Orange, CA, USA).

SOC medium

50 mg/mL apramycin (Apr)

LB+Apr plates

LB liquid medium

QIAgen Miniprep Kit

Restriction enzymes

38 mg/mL 2,6-diaminopimelic acid (DAP)

LB+Apr+DAP plates

R2-sucrose plates

Mixture of nalidixic acid and apramycin each at a concentration of 1 mg/mL

ISP+Apr

MYG liquid medium

Ethyl acetate

Heterologous expression of the spn cluster

1 Amplification of pathway fragments. The entire gene cluster is split into three intermediate plasmids, thus separating the 4 PKSs (Figure 4a). The first one

contains *spnDEFJAGKH*, the second one contains the last 400 bp of *spnH*, the full-length *spnA'* and the first 400 bp of *spnB*, and the third intermediate plasmid contains *spnBCILM*. Two restriction sites, *Ssp*I and *Pac*I are engineered at the appropriate positions of the intermediate plasmids by adding the corresponding recognition sites into the primers. Amplify the pathway fragments from the genomic DNA of *S. spectabilis* using the primers listed in Table 1. The genomic DNA can be isolated using Wizard® Genomic DNA Purification Kit. The following PCR components and conditions will be used as a standard protocol throughout the experiments. Set up the reaction mixtures as follows: 50 μL of FailSafe PCR 2×PreMix G, 2.5 μL of each primer (20 pmol/μL), 1 μL of template (at a concentration of 500–1000 ng/μL for genomic DNA and 10–50 ng/μL for plasmid), 1 μL of Phusion DNA polymerase, 5 μL of DMSO and 38 μL of ddH_2O in a total volume of 100 μL. Amplify pathway fragments using the following condition: Fully denature at 98 °C for 30 sec, followed by 25 cycles of 98 °C for 10 sec, 58 °C for 30 sec and 72 °C for 1–3 min, with a final extension at 72 °C for 10 min.

2 Amplification of helper fragments. Apparently, in the first step of assembly, the *S. lividans* helper fragment can be omitted. The *S. cerevisiae* helper fragment is amplified from the vector pRS416 (New England Biolabs). The *E. coli* helper fragment is amplified from pAE4, which is a *Streptomyces-E. coli* shuttle vector. The corresponding primers are listed in Table 1. Set up the reaction for amplifying the *S. cerevisiae* helper fragment as follows: 20 μL of 5 × Phusion GC Reaction buffer, 2.5 μL of dNTP premix, 2.5 μL of each primer (20 pmol/μL), 1 μL of template pRS416 at a concentration of 50–100 ng/μL, 1 μL of Phusion DNA polymerase, and 70.5 μL of ddH_2O in a total volume of 100 μL. Amplify the *S. cerevisiae* helper fragment using the following condition: Fully denature at 98 °C for 30 sec, followed by 25 cycles of 98 °C for 10 sec, 45 °C for 30 sec and 72 °C for 1 min, with a final extension at 72 °C for 10 min. Set up the reactions for amplifying the *E. coli* helper fragment as follows: 50 μL of FailSafe PCR 2 × PreMix G, 2.5 μL of each primer (20 pmol/μL), 1 μL of template plasmid pAE4 at a concentration of 50–100 ng/μL, 1 μL of Phusion DNA polymerase, and 43 μL of ddH_2O in a total volume of 100 μL. Use the standard reaction condition as described above.

We encountered difficulties in amplifying the S. cerevisaie helper fragment, and in the end, the correct product was only obtained by using the designated reagents and PCR program mentioned above. pAE4 was obtained from Professor William Metcalf at the University of Illinois, Urbana, IL, USA. The complete sequence of the plasmid can be obtained by request from the authors. As shown in the vector maps of the assembled construct (Fig. 4c), *the minimal genetic elements needed for selecting and maintaining successful transformants include origin of replication (ori) and selection marker. The S. cerevisiae helper fragment is designed to contain CEN6 and ARS H4 as an ori, and ura3 as a selection marker; the E. coli helper fragment is designed to contain oriR6K as an ori and accIV encoding for the apramycin resistance gene (it can be also*

used as the selection marker in Streptomyces). The Streptomyces helper fragment does not have the corresponding ori used in Streptomyces, therefore the assembled construct could not be maintained in the form of a plasmid, and instead has to be integrated to the chromosome. It also contains oriT which is the conjugal transfer ori, PhiC31 attP as the φC31 recognition site, int enconding for the φC31 integrase, and tL3 as a terminator.

3 Load 100 μL of PCR products onto 0.7% agarose gels and perform electrophoresis at 120 V for 20 minutes. Gel-purify PCR products using the QIAquick Gel Extraction Kit. Check the concentrations of the purified products using NanoDrop.

4 Take 200–300 ng of each fragment, mix in a tube, and calculate the final volume. Add 10% v/v 3 M sodium acetate and 2% v/v 10 mg/mL glycogen (e.g. if there is 100 μL of mixture, add 10 μL of sodium acetate and 2 μL of glycogen) and mix well. Add 2× v/v 100% ethanol (e.g. if the final volume is about 110 μL, add 220 μL ethanol) and mix well.

Store the DNA mixture at −80 °C for at least an hour. Centrifuge at 4 °C, 16,100 ×*g* for 20 minutes. Usually the precipitated DNA can be seen at the bottom of the tube. Remove the supernatant completely (do not touch the DNA). Add 500 μL of 70% ethanol to wash the DNA pellet, and centrifuge at room temperature, 16,100 ×*g* for 3 minutes. Remove the ethanol completely and air dry the pellet for 1–2 minutes (do not over-dry). Resuspend the DNA pellet with 4 μL of ddH_2O. Now the fragment mixture is ready for transformation. The fragment mixture can be maintained at −20 °C for several months.

Transformation

5 Inoculate a single colony of *S. cerevisiae* strain HZ848 into 3 mL of YPAD medium and grow overnight in a shaker at 30 °C and 250 rpm.

S. cerevisiae HZ848 (MATα, ade2-1, Δura3, his3–11, 15, trp1-1, leu2–3, 112, can1–100) is used as the host for DNA assembly. However, any S. cerevisiae auxotrophic strain can be used as a host if the vector carrying the corresponding selection marker is provided.

6 Measure the OD_{600} of the seed culture and inoculate the appropriate amount to 50 mL of fresh YPAD medium to obtain an OD_{600} of 0.2 (e.g. if the overnight culture has an OD_{600} of 10, then add 1 mL into 50 mL of fresh YPAD medium). The doubling time for a *S. cerevisiae* laboratory strain is approximately two hours. Continue growing the 50 mL of culture for approximately 4 hours to obtain an OD_{600} of 0.8.

7 Spin down the yeast cells at 4 °C, 4,000 rpm for 10 minutes and remove the spent medium. Use 50 mL of ice cold ddH_2O to wash the cells once and centrifuge again. Discard water, add 1 mL of ice-cold ddH_2O to resuspend the cells and move them to a sterile Eppendorf tube. Spin down the cells using a bench top centrifuge for 30 s at 4 °C, 7,000 rpm. Remove water and use 1 mL of

1 M ice-cold sorbitol to wash the cells once (now the cells look slightly yellow). Centrifuge again and remove the sorbitol. Resuspend the cells in 250–300 μL of chilled 1 M sorbitol and distribute them into 50 μL aliquots. Now each 50 μL of cells is ready for electroporation. Unlike *E. coli*, yeast competent cells need to be freshly prepared each time.

8 Mix 4 μL of DNA with 50 μL of yeast cells and put the mixture into a chilled electroporation cuvette. Electroporate the cells at 1.5 kV, and quickly add 1 mL of pre-warmed (30 °C) YPAD medium to resuspend cells. For an efficient electroporation, a time constant of 5.0–5.2 ms should be obtained. Grow in a shaker at 30 °C, 250 rpm for 1 hour. Spin down the cells in a sterile tube at 16,100 ×*g* for 30 s and remove the YPAD medium. Use 1 mL of room temperature sorbitol solution to wash the cells two to three times and finally resuspend the cells in 1 mL sorbitol. Spread 100 μL of resuspended cells onto SC-Ura plates. Incubate the plates at 30 °C for 2–3 days until colonies appear. Normally, 200–300 colonies will be obtained.

Verification of the correctly assembled constructs

9 Usually, ten colonies are randomly picked from the SC-Ura plate and each colony is inoculated into 1.5 mL of SC-Ura liquid medium. Grow at 30 °C for 1.5 days. Purify yeast plasmid DNA from each 1.5 mL of culture using the Zymoprep II kit. The assembly efficiency is defined as the percentage of the correct clones among the transformants appearing on the plate.

10 Mix 2 μL of isolated plasmid with 50 μL of *E. coli* BW25141 cells and put the mixture into a chilled electroporation cuvette. Electroporate the cells at 2.5 kV, and quickly add 1 mL of SOC medium to resuspend the cells. Grow in a shaker at 37 °C, 250 rpm for 1 hour. Spin down the cells, remove 800 μL of SOC medium, resuspend the pellet with the remaining 200 μL of SOC medium and spread the cells on LB+Apr plates. Incubate the plates at 37 °C for 16–18 hours until colonies appear.

E. coli strain BW25141 was used for plasmid enrichment and verification because we found this strain can stably maintain plasmids of large size. Any E. coli strains suitable for DNA cloning, such as DH5α and JM109, can be used for regular plasmids less than 20 kb. For an efficient electroporation, a time constant of 5.0–5.2 ms should be obtained. The number of obtained E. coli transformants could vary from a few to several thousands. This is mainly due to the low quality of the isolated yeast plasmids. However, as long as colonies appear, experiments can proceed.

11 Inoculate a single colony from each plate to 5 mL of LB supplemented with 50 μg/mL apramycin, and grow at 37 °C for 12–16 hours. Purify *E. coli* plasmids from each 5 mL of culture using the QIAgen Miniprep kit. Check the plasmid concentrations by NanoDrop.

12 Verify the correctly assembled pathway through restriction digestion. Set up the following digestions to verify the three intermediate constructs: (a) pZS-

SpnDEFJAGKH: *ApaL*I digestion at 37 °C for 3 hours: 1.5 μL of 10× NEB buffer, 0.15 μL of 100× BSA, 300 ng of plasmid, and 5 units of *ApaL*I. Add ddH_2O to a final volume of 15 μL. Expected bands: 609 bp, 778 bp, 2757 bp, 2853 bp, 3409 bp, 3786 bp, and 5450 bp. (b) pZS-Spn(H)A'(B): *Not*I digestion at 37 °C for 3 hours: 1.5 μL of 10× NEB buffer, 0.15 μL of 100× BSA, 300 ng of plasmid, and 5 units of *Not*I. Add ddH_2O to a final volume of 15 μL. Expected bands: 369 bp, 1335 bp, 2649 bp, 3576 bp, and 7160 bp. (c) pZS-SpnBCILM: *Sac*I digestion at 37 °C for 3 hours: 1.5 μL of 10× NEB buffer, 0.15 μL of 100× BSA, 300 ng of plasmid, and 5 units of *Sac*I. Add ddH_2O to a final volume of 15 μL. Expected bands: 752 bp, 939 bp, 2757 bp, 3086 bp, 3804 bp, 4091 bp, and 5992 bp.

In order to verify the correctly assembled constructs through restriction digestion, a set of digestion reactions consisting of one or two enzymes that cut the expected construct multiple times are chosen. Usually, for smaller constructs (<20 kb), find one enzyme that cuts the DNA molecule approximately 5–7 times and one set of digestion reactions is sufficient. Try to avoid using enzyme digestion that will result in multiple fragments with similar sizes. In addition to restriction digestion, the correctly assembled constructs can be confirmed by DNA sequencing.

The second-step assembly

13 Digest the confirmed intermediate plasmids with *Ssp*I and *Pac*I to generate three intermediate fragments (Figure 4b). Set up *Ssp*I/*Pac*I digestion at 37 °C for 3 hours: 5 μL of 10× NEB buffer, 1 μL of 100× BSA, 1 μg of plasmid, and 10 units of each enzyme. Add ddH_2O to a final volume of 50 μL. Gel-purify the 17,775 kb, 11,739 kb and 19,930 kb bands released from pZS-SpnDEFJAGKH, pZS-Spn(H)A'(B), and pZS-SpnBCILM, respectively, using the QIAquick Gel Extraction Kit. Check the concentrations by NanoDrop.

14 Mix 200–300 ng of intermediate fragments with a fragment obtained from restriction digestion of the master helper plasmid by *ApaL*I and *Xho*I. Concentrate the DNA mixture using the same procedures mentioned above.

The master helper plasmid was previously constructed by co-transforming the three PCR-amplified helper fragments, which shared overlaps of 40–80 bp, into S. cerevisiae. E, Y, and S represent the corresponding E. coli, S. cerevisiae, and Streptomyces helper fragments (Figure 4b).

15 Perform transformation and construct verification using the same procedures as for the intermediate constructs (step #5–12 under "Transformation" and "Verification of the correctly assembled constructs") Set up the following digestions to verify the final construct (Figure 4c): (a) *ApaL*I digestion at 37 °C for 3 hours: 1.5 μL of 10× NEB buffer, 0.15 μL of 100× BSA, 1 μg of plasmid, and 10 units of *ApaL*I. Add ddH_2O to a final volume of 15 μL. Expected bands: 609 bp, 778 bp, 813 bp, 2757 bp, 2853 bp, 3694 bp, 3786 bp, 3823 bp, 5734 bp, 8681 bp, 8756 bp, and 9656 bp. (b) *Sac*I digestion at 37 °C for 3 hours: 1.5 μL

of 10× NEB buffer, 0.15 μL of 100× BSA, 1 μg of plasmid, and 10 units of *Sac*I. Add ddH_2O to a final volume of 15 μL. Expected bands: 752 bp, 839 bp, 939 bp, 1806 bp, 1989 bp, 2757 bp, 2977 bp, 3804 bp, 5992 bp, 6560 bp, 6972 bp, 7937, and 8616 bp. For this step, an assembly efficiency of approximately 30% is expected.

For larger constructs (e.g. 40–50 kb), find one or two enzymes that cut the DNA molecule approximately 10 times and run two to three sets of digestion reactions in parallel to ensure the correct assembly.

Conjugation and heterologous expression of the spn pathway in *S. lividans*

16 Mix 50–100 ng of each verified plasmid with 50 μL of *E. coli* WM6026 cells and put the mixture into a chilled electroporation cuvette. Electroporate the cells at 2.5 kV, and quickly add 1 mL of SOC medium and 5 μL of 38 mg/mL DAP to resuspend cells. Grow in a shaker at 37 °C, 250 rpm for 1 hour. Spread 100 μL of each culture on a LB plate supplemented with apramycin and DAP. Incubate the plates at 37 °C for 16 hours until colonies appear.

E. coli strain WM6026 is obtained from Professor William Metcalf and serves as the donor strain for conjugating plasmids to S. lividans (the details of strain construction can be obtained by request from the authors). It is an auxotrophic strain whose growth relies on exogenously supplemented DAP.

17 Inoculate a single colony from each plate to 2 mL of LB supplemented with 50 μg/mL apramycin and 10 μL of 38 mg/mL DAP, and grow at 37 °C for approximately 2 hours until OD_{600} reaches 0.6~0.8. Spin down 100 μL of cell culture in an Eppendorf tube and wash the cell pellets each with 1 mL of fresh LB medium. Spin down the cells, and wash one more time. Resuspend the cell pellets each with 1 mL of LB. Mix 2 μL of the resuspended cells with 25 μL of *S. lividans* spores by pipetting and spot 2 μL of aliquots onto R2-sucrose plates. Wait until all the spotted drops are absorbed into the plates. Incubate the plates at 30 °C for 16–18 hours.

18 Flood the plates with 2 mL of a mixture of nalidixic acid and apramycin each at a concentration of 1 mg/mL. Incubate the plates at 30 $^{\circ}$C for additional 3–5 days until exconjugants appear, at which point exconjugants are picked and restreaked on ISP2-Apr plates and allowed to grow for 2 days.

Nalidixic acid is used to kill E. coli after it donates the plasmid into S. lividans, and apramycin is used to select the successful S. lividans exconjugants.

19 Inoculate single colonies from the ISP2-Apr plates into 10 mL of MYG supplemented with 50 μg/mL apramycin and grow the cultures at 30 °C for 2 days as seed cultures, of which 2.5 mL is subsequently inoculated to 250 mL of fresh MYG supplemented with 50 μg/mL apramycin and grown for another 84 hours.

Detection of spectinabilin

20 Centrifuge the cultures at 4,000 rpm for 10 minutes to remove the cells. Extract the supernatants with an equal volume of ethyl acetate and evaporate to dryness using a rotary evaporator.

21 Perform LC-MS on an Agilent 1100 series LC/MSD XCT plus ion trap mass spectrometer with an Agilent SB-C18 reverse-phase column. HPLC parameters for detection of spectinabilin are as follows: solvent A, 1% acetic acid in water; solvent B, acetonitrile; gradient, 10% B for 5 min, to 100% B in 10 min, maintain at 100% B for 5 min, return to 10% B in 10 min and finally maintain at 10% B for 7 min; flow rate 0.3 mL/min; detection by UV spectroscopy at 367 nm. Under such conditions, spectinabilin is eluted at 20.8 min. Mass spectra are acquired in ultra-scan mode using electrospray ionization (ESI) with positive polarity. The MS system is operated using a drying temperature of 350 $^{\circ}$C, a nebulizer pressure of 35 psi, a drying gas flow of 8.5 L min^{-1}, and a capillary voltage of 4500 V. The MS/MS pattern is identical to that of the authentic spectinabilin (Figure 5).

Gene inactivation and scar-less gene deletion

22 To inactivate *spnK*, change the codon TGC encoding Cys199 to a stop codon TGA in the PCR primers (Table 1 and Figure 1). Introducing a stop codon into the target gene located in the middle of the gene cluster instead of completely removing it has the advantage of maintaining the pathway structure without affecting the transcription of the neighboring genes. Remove the genes *spnM* and *spnM+spnL* individually from the biosynthetic pathway by redesigning the reverse primer for amplifying the last pathway fragment (Table 1). The assembly, verification, heterologous expression of the pathway variants, and detection of spectinabilin are performed using the same procedures as with the wild type gene cluster.

As a result, all three mutants will still produce spectinabilin (Figure 5), *indicating that spnK, spnL, and spnM are not required for spectinabilin biosynthesis, supporting our previous hypothesis that they are only involved in up-regulating the substrate concentration and product transportation* (Choi et al., 2010). *In addition to gene inactivation, site-directed mutagenesis can be readily used to create pathway variants. Unlike the labor-intensive and time-consuming procedures used in the conventional native host-based methods and heterologous host-based methods* (Blodgett et al., 2007; Ito et al., 2009; Karray et al., 2010), *DNA assembler only requires adding site-specific mutation(s) into the PCR primers used to generate pathway fragments.*

BASIC PROTOCOL 2. RECONSTRUCTING THE SILENT NOR GENE CLUSTER FROM *S. ORINOCI*

As mentioned above, when the direct cloning strategy was applied to the nor gene cluster from *S. orinoci*, no spectinabilin was detected. Unlike the spn pathway, the nor pathway

contains a repressor encoded by *norD*. Removing *norD* from the cluster did not lead to the production either, indicating that a more sophisticated regulation mechanism is embedded in spectinabilin biosynthesis in *S. orinoci*. The nor pathway therefore becomes a model pathway to test our cluster reconstructing strategy in activating silent gene clusters.

Our strategy is to select a single heterologous host, identify a set of promoters that are strong in the selected condition, assemble individual biosynthetic genes with these promoters into a new cluster, and express the reconstructed cluster in the selected condition (Figure 2). However, very few constitutive or inducible promoters have been characterized in Streptomycetes despite decades of research. In the following protocol, we will describe the procedures for identifying new promoters, confirming their activities, reconstructing the pathway and consequently activating the silent cluster. Note that this method was recently published (Shao et al., 2013).

Materials

In addition to materials listed in the basic protocol 1, the following materials are needed.

RNeasy Mini Kit (QIAGEN, Valencia, CA) containing buffer RLT, RW1, RPE) with ethanol supplemented

First-strand cDNA Synthesis Kit (Roche Applied Science, Indianapolis, IN)

SYBR® Green PCR Master Mix

TE buffer

Lysozyme

2-mecaptoethonol

Ambion® TURBO DNase Buffer, TURBO DNase and inactivation reagent (Life Technology, Carlsbad, CA)

0.1 M DTT

ε-caprolactam

Sample preparation

1 Select promoter candidates upstream of housekeeping genes originating from the final expression host or other closely related hosts.

We selected promoter candidates upstream of 23 housekeeping genes originated from Streptomyces griseus because its genome sequence is available, including RNA polymerase subunits, elongation factors, ribosomal proteins, glycolytic enzymes and various amino-acyl tRNA synthetases (Table 2).

2 Grow *S. griseus* in the fixed culture condition. Inoculate a single colony into 10 mL of MYG medium and grow in a shaker at 30 °C and 250 rpm until the culture reaches saturation, which usually takes 2 days. Inoculate 2 mL of the

saturated seed culture to 200 mL of fresh MYG and continue to grow in the same condition. Collect 10 mL of culture every 12 hours. Centrifuge the samples at 3,220 ×*g* for 10 min. Discard the supernatant. Store the pellets at −80 °C.

20–30 and 200–300 glass beads of 3–4 mm need to be added into the 10 mL medium and the 200 mL medium, respectively. The purpose is to break the growing mycelia in order to ensure their contact with enough oxygen. Streptomyces is a strictly aerobic organism. Limited air contact will severely affect cell growth. Here MYG is selected as the fixed culture condition because it is the most commonly used laboratory condition for culturing Streptomyces.

Total RNA extraction

The following protocol is designed based on the standard protocol provided in the RNeasy Mini Kit with a few modifications.

3 Extract the total RNA using the RNeasy Mini Kit. Thaw samples on ice and resuspend cells using 1 mL of TE buffer containing 5 mg of lysozyme. Shake the resuspended cells at 30 °C, 250 rpm for 20 min. All the tubes, tips and water should be RNase free.

4 Add 4 mL buffer RLT containing 400 μL of 2-mecaptoethonol. Mix completely by vortexing. Centrifuge at 3,220 ×*g* for 5 min at room temperature. Transfer supernatant to a new tube. Take out 600 μL of the supernatant for the next steps and store the rest at −80 °C.

5 Add 280 μL of 100% ethanol to the supernatant. Mix thoroughly by shaking. Do not centrifuge at this step. Transfer the sample to an RNeasy pink column, including any precipitate that may form. Spin at 16,100 ×*g* for 1 min.

6 Wash the column using 400 μL of buffer RW1. Spin at 16,100 ×*g* for 1 min. Wash the column using 250 μL of buffer RPE with ethanol supplemented. Spin at 16,100 ×*g* for 1 min. Wash the column again using 250 μL of buffer RPE with ethanol supplemented. Spin at 16,100 ×*g* for 1 min.

7 Transfer the pink column to a new collection tube. Spin at 16,100 ×*g* for 1 min to remove residual ethanol.

8 Transfer the pink column to a new collection tube. Elute with 22 μL of RNase-free water. Wait for 1 min and spin at 16,100 ×*g* for 1 min. The sample is ready for DNase treatment.

DNase treatment and cDNA synthesis

9 Add 2.5 μL of 10× TURBO DNase Buffer and 1 μL of TURBO DNase to the RNA, and mix gently. Incubate at 37 °C for 1 hour.

10 Add 3 μL of resuspended DNase inactivation reagent and pipette to mix well. Incubate at room temperature for 5 min and mix occasionally by gently pipetting. Centrifuge at 10,000 ×*g* for 1.5 min and transfer the supernatant to a fresh tube.

The DNase inactivation reagent contains magnetic beads. Occasional mixing will ensure the complete removal of DNase. When transferring supernatant, try to avoid touching the beads. The residual beads will interfere with the following reaction. If needed, repeat the centrifuge step and transfer the supernatant to a fresh tube.

11 Before performing reverse transcription, mix the following reaction components: 10 μL of RNA, 1 μL of oligo dT primer or random hexamer (served with the kit), 1 μL of RNase-free water, 1 μL of 10 mM dNTP. Incubate at 65 °C for 5 min and immediately put on ice for at least 1 min. This step ensures denaturation of RNA secondary structures. Split the reaction mixture into halves.

12 Perform reverse transcription using the First-strand cDNA Synthesis Kit. In one half of the above mixture, add the following reagents: 4 μL of 5× first-strand buffer, 1 μL of 0.1M DTT, 1 μL of RNase OUT recombinant RNase inhibitor, 1 μL of SuperScript III reverse transcriptase (200 U/μL). The second half of the mixture obtained in the last step will work as a negative control. Add everything else except the reverse transcriptase and replace its volume by 1 μL of H_2O. Mix gently and perform the cDNA synthesis using the following condition: 25 °C for 5 min, followed by 50 °C for 30 min, and 70 °C for 15 min, with a final holding temperature at 4 °C. Now the synthesized cDNA is ready for real-time PCR.

The negative reaction is used for monitoring the removal of genomic DNA in the DNase digestion step. The residual genomic DNA will give significant amplicons in real-time PCR.

Real-time PCR

13 According to the sequences of the 23 candidate housekeeping genes from *S. griseus* (Table 2), primers for real-time PCR are designed and synthesized by Integrated DNA Technologies (Coralville, Iowa). The expression of *hrdB* encoding for RNA polymerase sigma factor is used as the internal control, and the expression of other candidate genes are normalized by that control.

Other Streptomyces species whose genome sequences are available can also be an appropriate source for the promoter screening. HrdB is a housekeeping sigma factor commonly used as the internal control for transcription analysis in Streptomyces (Gomez et al., 2011; Kim et al., 2011; Nieselt et al., 2010).

14 Prepare genomic DNA samples for constructing standard curves. Before setting up real-time PCR, prepare a serial dilution of *S. griseus* genomic DNA, which is isolated following the same procedure as with *S. spectabilis* and *S. orinoci*. Measure the DNA concentration by Nanodrop and dilute the genomic DNA to concentrations of 64 ng/μL, 4 ng/μL, 0.25 ng/μL, 0.015625 ng/μL (16× serial dilution).

Genomic DNA is used for constructing standard curves because it provides a known ratio between the target genes and the internal control.

15 Set up the real-time PCR reactions as follows: 10 μL of SYBR® Green PCR Master Mix, 0.5 μL of forward primer (20 pmol/μL), 0.5 μL of reverse primer (20 pmol/μL), 1 μL of template, and 8 μL of H_2O to make up a final volume to 20 μL. Templates include cDNA sample, the negative control reaction product in the cDNA synthesis step, and the genomic DNA samples of different concentrations. Pipette the reaction mixtures into individual well on a MicroAmp® optical 384-well reaction plate (Applied Biosystems, Carlsbad, CA).

16 Perform the reactions on the 7900HT Fast Real-Time PCR System (Applied Biosystems) using the following condition: 50 °C for 2 min, 95 °C for 10 min, 40 cycles of 95 °C for 15 sec and 60 °C for 1 min, followed by a final dissociation stage at 95 °C for 15 sec, 60 °C for 15 sec and 95 °C for 15 sec. Use an appropriate detector for SYBR green detection. Analyze data by the software SDS2.4 provided by Applied Biosystem. Calculate the concentrations of cDNA using individual standard curves. The results are expressed in forms of ng/μL equivalent to genomic DNA. Normalize the expression by the calculated cDNA concentration of the internal control.

The expected result: two genes encoding glyceraldehyde-3-phosphate dehydrogenase (gapdh) and 30S ribosomal protein S12 (rpsL) should stand out with much higher transcription levels than all the other genes in the fixed culturing condition (Figure 6a), *indicating that their corresponding promoters could be very strong. The gapdh promoter, named as* $gapdh_p$ *(SG), is located upstream of the gapdh operon consisting of Gapdh, phosphoglycerate kinase (pgk) and triosephosphate isomerase (tpiA), the enzymes catalyzing the* 6^{th}, 7^{th} *and* 5^{th} *steps, respectively, in the glycolysis pathway; and the rpsL promoter, named as* $rpsL_p$ *(SG), resides upstream of another operon consisting of 30s ribosomal proteins S12 and S7, and elongation factor G and Tu.*

Confirming promoter activity

17 Clone these two candidate promoters upstream of the Streptomycete reporter gene, *xylE*, encoding catechol 2, 3-dioxygenase. Previously, *xylE* has been cloned downstream of the promoter, $ermE^*_p$, in pAE4 vector, giving the construct pAE4-ermE*p-xylE. PCR amplify the promoter from the *S. griseus* genomic DNA, and the first 262 bp sequence of *xylE* (before the *Nde*I site, named as xylE262) from pAE4-ermE*p-xylE. Restriction sites *Afl*II and *Nde*I are incorporated into the primers (Table 3). Set up PCR reactions using the standard protocol without DMSO. Perform electrophoresis, gel-purify PCR products, and check the concentrations using the procedures described above.

The entire intergenic region between the target gene and its upstream gene is cloned upstream of xylE. Here we do not provide additional information to experimentally determine the ribosomal binding site for each promoter and assume it is located 6–10 bp upstream of each start codon. However, we did find these intergenic regions are AG-rich for most promoters.

The promoter ermE$_p^*$*, is the mutated variant of the promoter of the erythromycin resistance gene from Saccharopolyspora erythraea, and is believed to be one of the strongest constitutive promoters in Streptomycetes* (Schmitt-John and Engels, 1992; Wagner et al., 2009).

18 For overlap extension PCR, set up the first-step reaction mixture as follows: 10 μL of FailSafe PCR 2× PreMix G, 100 ng of promoter fragment, 100 ng of xylE262 fragment, and 0.2 μL of DNA polymerase. Add ddH_2O to a final volume of 20 μL. Reaction condition: fully denature at 98 °C for 30 sec, followed by 10 cycles of 98 °C for 10 sec, 58 °C for 30 sec and 72 °C for 1 min, with a final extension at 72 °C for 10 min. Set up the second-step reaction mixture as follows: 50 μL of FailSafe PCR 2× PreMix G, 10 μL of first-step reaction mixture, 1 μL of DNA polymerase, 2.5 μL of forward primer (20 pmol/μL), 2.5 μL of reverse primer (20 pmol/μL), and 34 μL ddH_2O in a total volume of 100 μL. Run the reactions using the standard PCR protocol. Perform electrophoresis, gel-purify PCR products, and check the concentrations using the procedures described above.

19 Digest pAE4-ermE*p-xylE and the OE-PCR products by *Afl*II and *Nde*I at 37 °C for 3 hours. Digestion condition: 5 μL of 10× buffer, 1 μL of 100× BSA, 1 μg of OE-PCR product or 3 μg of pAE4-ermE*p-xylE, and 10 units of each enzyme per μg of DNA. Add ddH_2O to a final volume of 50 μL. Load the digestion products onto 0.7% agarose gels and perform electrophoresis at 120 V for 20–30 minutes. Gel-purify the digested OE-PCR product and the 5,545 kb band from pAE4-ermE*p-xylE digestion using the QIAquick Gel Extraction Kit and check the concentrations of the purified products using NanoDrop.

22 Set up the ligation mixture as follows: 2 μL of 10× T4 DNA ligation buffer, 200 ng of the digested pAE4-ermE*p-xylE backbone, 200 ng of the digested OE-PCR product, 1 μL of T4 DNA ligase (400, 000 Units/μL). Add ddH_2O to a final volume of 20 μL. Incubate at 16 °C for 12–16 hours.

23 Mix 2 μL of the ligation mixture with 50 μL of *E. coli* BW25141 cells and put the mixture into a chilled electroporation cuvette. Electroporate the cells at 2.5 kV, and quickly add 1 mL of SOC medium to resuspend the cells. Grow in a shaker at 37 °C, 250 rpm for 1 hour. Spin down the cells, remove 800 μL of SOC medium, resuspend the pellet with the remaining 200 μL of SOC medium and spread the entire volume on LB+Apr plates. Incubate the plates at 37 °C for 16–18 hours until colonies appear.

24 Inoculate a single colony from each plate to 5 mL of LB+Apr, and grow at 37 °C for 12–16 hours. Purify *E. coli* plasmids from each 5 mL of culture using the QIAgen Miniprep kit. Check the plasmid concentrations by NanoDrop. Verify the correct clones through restriction digestion by *Afl*II and *Nde*I.

25 Transform and conjugate the verified constructs into *S. lividans* using the same procedures described above for the spn pathway. Apply the same condition used to prepare samples in promoter screening to express *xylE*. Collect samples at 24

hours and 48 hours and perform a XylE activity assay as described elsewhere (Tobias Kieser and Hopwood, 2000). The same experiments are performed on pAE4-ermE*p-xylE, which serves as a control.

A figure illustrating the expected result is shown in Figure 6b. *Compared to the control promoter, much stronger activities of gapdh$_p$ (SG) and rpsL$_p$ (SG) should be observed. The strong activities of gapdh$_p$ and the promoters of various translation elongation factors have also been observed in many other microorganisms, such as fungi, bacteria, micro-algae, and protozoa* (Ahmad et al., 2012; Ahn et al., 2007; Hong et al., 2010; Jia et al., 2012; Krasny et al., 2000; Shao et al., 2009; Suarez et al., 2006).

Cloning additional promoters

26 Clone the corresponding gapdh$_p$ and rpsL$_p$ from other actinobacteria. Their sequences may be downloaded from the NCBI website. Here, we choose 18 distinct actinobacteria (Table 4) and clone the corresponding promoters upstream of *xylE*.

Encouraged by the strong activities of gapdh$_p$ and rpsL$_p$, we could examine their equivalents from other species. The promoters from different Streptomyces species share extremely high homologies with each other, which is undesired because such high sequence similarities would cause severe deletions during DNA assembly in S. cerevisiae. Since Streptomycetes belong to actinobacteria family, the promoters from other genera of actinobacteria could possibly be recognized by the transcription machinery in Streptomycetes as well. The gapdh and the rpsL operon are highly conserved in the actinobacteria family, but the corresponding promoters are highly diversified.

27 Obtain promoters from these hosts through primer splicing, in which 6–10 overlapping oligonucleotides designed based on the sequence of each promoter are joined through overlap extension PCR (Figure 7). Set up the first-step reaction mixture as follows: 10 μL of FailSafe PCR 2× PreMix G, 0.5 μL of each primer (20 pmol/μL), and 0.2 μL of DNA polymerase. Add ddH_2O to a final volume of 20 μL. Reaction condition: fully denature at 98 °C for 30 sec, followed by 10 cycles of 98 °C for 10 sec, 58 °C for 30 sec and 72 °C for 1 min, with a final extension at 72 °C for 10 min. Set up the second-step reaction mixture as follows: 50 μL of FailSafe PCR 2× PreMix G, 10 μL of first-step reaction mixture, 1 μL of DNA polymerase, 2.5 μL of the first primer and the last primer (20 pmol/μL), 2.5 μL of reverse primer (20 pmol/μL), and 34 μL ddH_2O in a total volume of 100 μL. Run the reactions using the standard PCR protocol. Perform electrophoresis, gel-purify PCR products, and check the concentrations using the procedures described above.

28 Splice the promoters individually with the fragment xylE262 *via* overlap extension PCR, clone the resulting product into pAE4, perform XylE assay using the same procedures as with gapdh$_p$ (SG) and rpsL$_p$ (SG).

As a result, 13 out of the 36 promoter candidates should be shown to be very active in S. lividans, many of them having more than 10-fold higher activities than ermE$_p$* (Figure 6b).

Reconstructing the nor cluster

29 Pick nine strong constitutive promoters from the collection to drive the expression of the nor pathway genes except for *norD* and *norG*. NorD is the repressor and NorG is the first enzyme in the pathway, converting chorismate from the shikimate pathway to *p*-aminobenzoic acid. The expression of *norG* will be driven by the ε-caprolactam inducible promoter, nitA$_p$ (Herai et al., 2004) (Figure 8a).

30 Use the previously mentioned 2-step assembly strategy to reconstruct the 42.6 kb nor pathway. Design primers to give overlaps of 80 bps between promoters and genes (Figure 3b and Table 5). Amplify the promoters from the corresponding pAE-promoter-xylE plasmids and amplify pathway genes from the genomic DNA of *S. orinoci*. Similar to the first step in the assembly of the spn pathway, the first-step assembly will result in three intermediate plasmids (Figure 8b). Perform restriction digestion on the verified constructs by *Avr*II and *Ssp*I to release intermediate pathway fragments, which will subsequently be combined with the master helper fragment as a full-length nor pathway in the second-step assembly. The transformation and verification protocols are the same with those for the spn pathway.

Expressing the reconstructed nor cluster and detecting the product

31 Follow the above mentioned protocol to conjugate the assembled final construct to *S. lividans*. Pick the resulting exconjugants and restreak on ISP2+Apr plates and allow growing for 2 days. Inoculate a single colony into 10 mL of MYG medium supplemented with 50 μg/mL apramycin and grow at 30 °C for 2 days as a seed culture, of which 2.5 mL will be subsequently inoculated to 250 mL of fresh MYG medium. Shake the culture at 30 °C, 250 rpm for 5 days. Add the inducer, ε-caprolactam, to the culture at a concentration of 1 g/L after 12 hours, and collect samples at appropriate times afterward. Follow the same protocol for spectinabilin extraction and detection.

As an expected result, the reconstructed pathway should be successfully activated, with a titer of approximately 0.1 mg/L (Figure 9).

REAGENTS AND SOLUTIONS

YPAD

Dissolve 6 g yeast extract, 12 g peptone, 12 g dextrose, and 60 mg adenine hemisulphate in 600 mL of ddH_2O. Autoclave at 121 °C for 15 min.

SC-Ura liquid medium and agar plates (synthetic complete drop-out medium lacking uracil)

Dissolve 3 g ammonium sulfate, 1 g yeast nitrogen source without ammonium sulfate and amino acids, 0.5 g complete synthetic medium minus uracil (CSM-Ura) purchased from MP Biomedicals (Solon, OH), 26 mg adenine hemisulphate, and 12 g dextrose in 600 mL of ddH_2O and adjust pH to 5.6 by NaOH. Autoclave at 121 °C for 15 min. To make SC-Ura plates, add 12 g agar before autoclaving.

SOC

Add 20 g Bacto-tryptone, 5 g yeast extract, 0.5 g NaCl, 186.4 mg KCl into 980 mL of ddH_2O. Adjust pH to 7.0 with NaOH. Autoclave at 121 °C for 15 min. After the solution cools down to 70–80 °C, add 20 mL of sterile 1 M glucose.

LB liquid medium and LB+Apr agar plates

Add 10 g bacto-tryptone, 5 g yeast extract, 10 g NaCl into 1 L of ddH_2O. Autoclave at 121 °C for 15 min. To make agar plates, add 20 g agar into the solution before autoclaving. After the solution cools down to 70–80 °C, add 1 mL of 50 mg/mL apramycin. Pour 20–25 mL into each petri dish.

MYG liquid medium

Add 10 g malt extract, 4 g yeast extract, and 4 g glucose into 1 L of ddH_2O. Adjust pH to 7.2 with NaOH. Autoclave at 121 °C for 15 min.

ISP2+Apr agar plates

Add 10 g malt extract, 4 g yeast extract, 4 g glucose, and 20 g agar into 1 L of ddH_2O. Adjust pH to 7.2 with NaOH. Autoclave at 121 °C for 15 min. After the solution cools down to 70–80 °C, add 1 mL of 50 mg/mL apramycin. Pour 20–25 mL into each petri dish.

R2-sucrose agar plates

Add 0.25 g of K_2SO_4, 10.12 g of $MgCl_2 \cdot 6H_2O$, 10 g of glucose, 0.1 g of Difco casamino acids, 5.73 g of TES into 990 mL of deionized water. Autoclave at 121 °C for 15 min. After the solution cools down to 70–80 °C, add 1 mL of 5 mg/mL KH_2PO_4, 8 mL of 36.8 mg/mL $CaCl_2 \cdot 2H_2O$, 1.5 mL of 0.2 g/mL L-proline and 0.5 mL of 1M NaOH (All these solutions need to be filter-sterilized in advance). Pour 20–25 mL into each Petri dish. Please refer to the recipe in the book “Practical Streptomyces Genetics”, page 408 (Tobias Kieser and Hopwood, 2000).

TE buffer

10 mM Tris, 1 mM EDTA, adjust pH to 8.0 with HCl.

COMMENTARY

Background Information

Microorganisms and plants produce numerous secondary metabolites or natural products, which are a prolific source of therapeutic agents (Dewick, 2002; Herbert, 1989; Li and Vederas, 2009). For example, it has been estimated that 77% of antibacterial drugs and 78% of anticancer drugs are either natural products or their derivatives (Newman and Cragg, 2012). Existing strategies for characterizing natural product gene clusters can be broadly classified into two groups including the native host gene inactivation method and the heterologous host gene expression method (Shao et al., 2011). The native host-based method involves introducing mutation(s) and deleting gene(s) directly from the chromosome and comparing the metabolite profile of the mutant with that of the wild type. In comparison, the heterologous host-based method involves moving the gene cluster from the native producer to a secondary host and comparing the metabolite profile of the resulting strain with that of the background strain. However, both approaches suffer a few limitations. For the native host-based method, firstly, most native producers are not cultivable in the laboratory and many microorganisms grow very slowly and produce minute amount of target compounds. It has been estimated that only 1% of bacteria and 5% of fungi have been cultivated in the laboratory (Bull et al., 1992; Davies, 1999; Demain, 2006; Leadbetter, 2003). Secondly, genetic tools have not been developed for the majority of organisms, which makes manipulation of a target biosynthetic pathway inconvenient and inefficient since each host has to be examined individually. On the other hand, the heterologous host-based method suffers from difficulties in the transfer and genetic manipulation of large biosynthetic pathways (>30 kb), problems with promoter recognition and gene expression, the limited availability of expression hosts, and low yield (Gross, 2007). Therefore, it is highly desirable to establish novel efficient modification, transfer, and expression technologies.

Moreover, recent advances in DNA sequencing technologies and bioinformatics indicate that sequenced genomes and metagenomes may represent a tremendously rich source for discovery of novel pathways involved in natural product biosynthesis (Challis, 2008; Zerikly and Challis, 2009). However, only a tiny fraction of these biosynthetic pathways have been characterized. For example, genome mining of Streptomyces (Ohnishi et al., 2008; Omura et al., 2001), myxobacteria (Bode and Muller, 2006), cyanobacteria (Abed et al., 2009; Singh et al., 2005) and fungi (Brakhage and Schroeckh, 2011) revealed the presence of many cryptic pathways whose products have not been identified while these strains were previously known to produce only a few compounds before their genomes were sequenced. The biosynthesis of natural products is highly regulated and the regulation is conducted through dozens of pleiotropic regulatory genes and pathway-specific regulators (Bibb and Hesketh, 2009; Bibb, 2005; Brakhage et al., 2008; van Wezel et al., 2009). They interact with each other to form an extremely complex network in response to a variety of physiological and environmental signals. These embedded complex regulations always hamper the discovery of novel natural products. Existing approaches to elicit pathway expression mainly include: (1) manipulating cell culture parameters, such as medium composition, to ensure expression of pathway-specific activator(s), presence of physiological and environmental co-inducer(s), or derepression of the genes repressed by

repressor(s) (Brakhage et al., 2008; van Wezel et al., 2009); (2) engineering the regulation by expressing the pathway-specific regulator under a well-characterized promoter (Bergmann et al., 2007; Wendt-Pienkowski et al., 2005); (3) testing a variety of heterologous hosts to express the target cluster (Baltz, 2010); (4) silencing major secondary metabolite biosynthetic pathways in order to simplify product identification and relieve competition for key precursors (Komatsu et al., 2010); (5) utilizing industrial strains that have been set up for high-level production of specific compounds (Rodriguez et al., 2003). Obviously, all these strategies can only be applied on a case-by-case basis. Each gene cluster has its own unique regulatory mechanism that has to be examined individually in order to identify the suitable context for the cluster to be activated. Such methods suffer from the laborious process, lack general applicability, and repeatedly identify compounds that are already known.

The strategy we present here remedies several of the key drawbacks associated with the traditional native host-based method and the heterologous host-based method, offering a powerful way to study natural product biosynthetic clusters and shedding light on deciphering cryptic pathways identified from genome mining (Challis, 2008; Gross, 2007; Zerikly and Challis, 2009).

Critical Parameters

The direct cloning strategy is used mostly when the cluster has been demonstrated to be functional in the heterologous host, and the product has been identified, but the biosynthetic mechanism has not been fully clarified. This strategy bypasses the conventional time-consuming processes to perform gene disruption and could easily be used to introduce fine modification without any remaining markers. On the other hand, the cluster reconstructing strategy is more generic and can be applied when the cluster in the background of native regulation is not guaranteed in the heterologous host. When the product from the reconstructed pathway has been successfully identified, the concepts of introducing gene disruption and fine-modification in the direct cloning strategy can be easily applied. The key steps in the cluster reconstructing strategy include: (1) Determine an ideal expression host, which in general should be closely related to the native producer in order to provide both necessary precursors and a similar environment for protein translation and folding. The choice of the expression host should also be made based on the availability of genetic tools to manipulate the organism. For example, *S. lividans* was used as a host to express clusters from other *Streptomyces* species (Eliot et al., 2008; Felnagle et al., 2007; Feng et al., 2009; Tobias Kieser, 2000; Woodyer et al., 2006) and *S. cerevisiae* was used to express clusters from fungi (Ishiuchi et al., 2012; Wawrzyn et al., 2012). (2) Identify a set of strong constitutive promoters. Except for *E. coli* and *S. cerevisiae*, most organisms do not have such a set of promoters reported in literature for ready use. Thanks to the development of real-time PCR and RNA-seq, strong constitutive promoters can be identified rapidly from the upstream regions of the housekeeping genes in the selected expression host. The culturing condition for promoter identification will subsequently be used to produce the target compound such that the transcription of each pathway gene is forced to be turned on. As we demonstrate here, promoters from closely-related organisms have a very high chance to be recognized by the transcription machinery in the selected host. Note that these

promoters could undergo unknown regulations in the target host, thus might not be completely constitutive. But at least expression of the downstream genes should consistently be observed. (3) PCR-amplify pathway genes from the target cluster using primers that will generate an overlap between adjacent fragments and refactor the cluster using DNA assembler in *S. cerevisiae*. Several other DNA assembly methods such as Gibson cloning (Gibson et al., 2009), RecET mediated direct cloning (Cobb and Zhao, 2012; Fu et al., 2012), and reiterative recombination (Wingler and Cornish, 2011), which are all based on homologous recombination, were reported recently. DNA assembler is known for its easy protocol and high assembly accuracy for large constructs by taking advantage of yeast *in vivo* homology recombination. Pathways up to 50 kb can be assembled routinely within 1–2 week with an efficiency of 30–100% (Shao et al., 2011). Such an *in vivo* homologous recombination-based assembly has also been used to assemble molecules as large as a genome (Gibson et al., 2008). Combined with the rapid promoter identification mentioned above, the method we propose here enables pathway design and manipulation in any desired organism as easily as in *E. coli* and *S. cerevisiae*. Codons can be optimized at this step if needed. (4) Express the cluster in the selected host and identify the product. Metabolites extracted from the expression host carrying the refactored cluster are usually analyzed by LC and compared with those extracted from the host lacking the exogenous pathway or carrying a non-functional pathway (e.g. a pathway with an essential gene deleted or mutated). Compounds appearing as new peaks are purified and subjected to mass spectrometry or NMR analysis for structure clarification.

Troubleshooting

PCR—In some cases, especially for amplifying certain fragments from Streptomyces, fine-tuning of the annealing temperature is often necessary for obtaining the correct amplicons or improving PCR yields. Due to the high GC content of Streptomyces genome (>70%), including 5% DMSO in the reaction mixture will reduce the chance of forming secondary structures in general, resulting in better amplification efficiency. However, in some cases, we did encounter lower amplification efficiencies or even failures when DMSO was used. 58 °C is used as a standard annealing temperature. Generally, when difficulties are encountered to amplify a certain fragment, a set of PCR conditions with various annealing temperatures and inclusion or exclusion of 5% of DMSO need to be tested.

Real-time PCR—Primers for real-time PCR are designed according to the online software (http://www.idtdna.com/scitools/Applications/RealTimePCR/) provided by Integrated DNA Technologies (Coralville, Iowa). For each reaction, a dissociation curve should be examined to ensure the reaction quality. Nonspecific amplification will result in more than one major peak, in which case, new primers need to be attempted.

Pathway assembly—The failure in using one-step assembly to construct the spn pathway was not due to the large size of the cluster because clusters of similar sizes (~ 50 kb) can be routinely obtained in our laboratory. Whether high sequence similarity between genes could potentially be problematic in assembly varies case-by-case. It did not seem to be a problem in the assembly of the aureothin gene cluster which has three PKSs (Shao et al., 2011). The one-step assembly works efficiently for most gene clusters lacking highly repeated

sequences. When handling clusters containing highly repeated sequences, we recommend attempting the one-step assembly first, and if it fails, then the cluster should be split into a couple of smaller intermediate plasmids as described above.

Anticipated Results

For a construct of less than 20 kb, an assembly efficiency of 70–100% is expected, and for a construct of 20–50 kb, an assembly efficiency of at least 30% is expected.

Time Considerations

The corresponding time required by each major step is listed in Table 6.

Acknowledgments

This work was supported by the National academies Keck Futures Initiative on Synthetic Biology and the National Institutes of Health (GM077596).

Literature Cited

Abed RM, Dobretsov S, Sudesh K. Applications of cyanobacteria in biotechnology. J Appl Microbiol. 2009; 106:1–12. [PubMed: 19191979]

Ahmad I, Shim WY, Jeon WY, Yoon BH, Kim JH. Enhancement of xylitol production in Candida tropicalis by co-expression of two genes involved in pentose phosphate pathway. Bioprocess and Biosystems Engineering. 2012; 35:199–204. [PubMed: 21969058]

Ahn J, Hong J, Lee H, Park M, Lee E, Kim C, Choi E, Jung J, Lee H. Translation elongation factor 1-alpha gene from Pichia pastoris: molecular cloning, sequence, and use of its promoter. Applied Microbiology and Biotechnology. 2007; 74:601–608. [PubMed: 17124582]

Baltz RH. Streptomyces and Saccharopolyspora hosts for heterologous expression of secondary metabolite gene clusters. J Ind Microbiol Biotechnol. 2010; 37:759–772. [PubMed: 20467781]

Bergmann S, Schumann J, Scherlach K, Lange C, Brakhage AA, Hertweck C. Genomics-driven discovery of PKS-NRPS hybrid metabolites from Aspergillus nidulans. Nat Chem Biol. 2007; 3:213–217. [PubMed: 17369821]

Bibb M, Hesketh A. Chapter 4. Analyzing the regulation of antibiotic production in streptomycetes. Methods Enzymol. 2009; 458:93–116. [PubMed: 19374980]

Bibb MJ. Regulation of secondary metabolism in streptomycetes. Current Opinion in Microbiology. 2005; 8:208–215. [PubMed: 15802254]

Blodgett JA, Thomas PM, Li G, Velasquez JE, van der Donk WA, Kelleher NL, Metcalf WW. Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide. Nat Chem Biol. 2007; 3:480–485. [PubMed: 17632514]

Bode HB, Muller R. Analysis of myxobacterial secondary metabolism goes molecular. J Ind Microbiol Biotechnol. 2006; 33:577–588. [PubMed: 16491362]

Brakhage AA, Schroeckh V. Fungal secondary metabolites - strategies to activate silent gene clusters. Fungal Genet Biol. 2011; 48:15–22. [PubMed: 20433937]

Brakhage AA, Schuemann J, Bergmann S, Scherlach K, Schroeckh V, Hertweck C. Activation of fungal silent gene clusters: a new avenue to drug discovery. Prog Drug Res. 2008; 66(1):3–12.

Bull AT, Goodfellow M, Slater JH. Biodiversity as a source of innovation in biotechnology. Annu Rev Microbiol. 1992; 46:219–252. [PubMed: 1444255]

Challis GL. Mining microbial genomes for new natural products and biosynthetic pathways. Microbiology-Sgm. 2008; 154:1555–1569.

Choi YS, Johannes TW, Simurdiak M, Shao Z, Lu H, Zhao H. Cloning and heterologous expression of the spectinabilin biosynthetic gene cluster from *Streptomyces spectabilis*. Mol Biosyst. 2010; 6:336–338. [PubMed: 20094652]

Cobb RE, Zhao H. Direct cloning of large genomic sequences. Nat Biotechnol. 2012; 30:405–406. [PubMed: 22565964]

Davies J. Millennium bugs. Trends Cell Biol. 1999; 9:M2–M5. [PubMed: 10611670]

Demain AL. From natural products discovery to commercialization: a success story. J Ind Microbiol Biotechnol. 2006; 33:486–495. [PubMed: 16402247]

Dewick, PM. A biosynthetic approach. 2nd ed.. Chichester, UK: John Wiley and Sons; 2002. Medical natural products.

Eliot AC, Griffin BM, Thomas PM, Johannes TW, Kelleher NL, Zhao H, Metcalf WW. Cloning, expression, and biochemical characterization of *Streptomyces rubellomurinus* genes required for biosynthesis of antimalarial compound FR900098. Chem Biol. 2008; 15:765–770. [PubMed: 18721747]

Felnagle EA, Rondon MR, Berti AD, Crosby HA, Thomas MG. Identification of the biosynthetic gene cluster and an additional gene for resistance to the antituberculosis drug capreomycin. Applied and Environmental Microbiology. 2007; 73:4162–4170. [PubMed: 17496129]

Feng Z, Wang L, Rajski SR, Xu Z, Coeffet-LeGal MF, Shen B. Engineered production of iso-migrastatin in heterologous Streptomyces hosts. Bioorg Med Chem. 2009; 17:2147–2153. [PubMed: 19010685]

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; 30:440–446. [PubMed: 22544021]

Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, Glass JI, Venter JC, Hutchison CA, Smith HO. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science. 2008; 319:1215–1220. [PubMed: 18218864]

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods. 2009; 6 343-U341.

Gomez C, Horna DH, Olano C, Palomino-Schatzlein M, Pineda-Lucena A, Carbajo RJ, Brana AF, Mendez C, Salas JA. Amino acid precursor supply in the biosynthesis of the RNA polymerase inhibitor streptolydigin by Streptomyces lydicus. J Bacteriol. 2011; 193:4214–4223. [PubMed: 21665968]

Gross H. Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. Applied Microbiology and Biotechnology. 2007; 75:267–277. [PubMed: 17340107]

Herai S, Hashimoto Y, Higashibata H, Maseda H, Ikeda H, Omura S, Kobayashi M. Hyper-inducible expression system for streptomycetes. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101:14031–14035. [PubMed: 15377796]

Herbert, RB. The biosynthesis of secondary metabolites. 2nd ed.. London, UK: Chapman and Hall; 1989.

Hong WK, Kim CH, Heo SY, Luo L, Oh BR, Seo JW. Enhanced production of ethanol from glycerol by engineered Hansenula polymorpha expressing pyruvate decarboxylase and aldehyde dehydrogenase genes from Zymomonas mobilis. Biotechnology Letters. 2010; 32:1077–1082. [PubMed: 20354759]

Isaka M, Jaturapat A, Kramyu J, Tanticharoen M, Thebtaranonth Y. Potent *in vitro* antimalarial activity of metacycloprodigiosin isolated from *Streptomyces spectabilis* BCC 4785. Antimicrob Agents Chemother. 2002; 46:1112–1113. [PubMed: 11897600]

Ishiuchi K, Nakazawa T, Ookuma T, Sugimoto S, Sato M, Tsunematsu Y, Ishikawa N, Noguchi H, Hotta K, Moriya H, Watanabe K. Establishing a new methodology for genome mining and biosynthesis of polyketides and peptides through yeast molecular genetics. Chembiochem. 2012; 13:846–854. [PubMed: 22447505]

Ito T, Roongsawang N, Shirasaka N, Lu W, Flatt PM, Kasanah N, Miranda C, Mahmud T. Deciphering pactamycin biosynthesis and engineered production of new pactamycin analogues. Chembiochem. 2009; 10:2253–2265. [PubMed: 19670201]

Jia YL, Li SK, Allen G, Feng SY, Xue LX. A novel glyceraldehyde-3-phosphate dehydrogenase (GAPDH) promoter for expressing transgenes in the halotolerant alga Dunaliella salina. Current Microbiology. 2012; 64:506–513. [PubMed: 22371187]

Kakinuma K, Hanson CA, Rinehart KL Jr. Spectinabilin, a new nitro-containing metabolite isolated from streptomyces spectabilis. Tetrahedron. 1976; 32:217–222.

Karray F, Darbon E, Nguyen HC, Gagnat J, Pernodet JL. Regulation of the biosynthesis of the macrolide antibiotic spiramycin in Streptomyces ambofaciens. J Bacteriol. 2010; 192:5813–5821. [PubMed: 20817767]

Kim SH, Lee HN, Kim HJ, Kim ES. Transcriptome analysis of an antibiotic downregulator mutant and synergistic Actinorhodin stimulation via disruption of a precursor flux regulator in Streptomyces coelicolor. Appl Environ Microbiol. 2011; 77:1872–1877. [PubMed: 21216912]

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 U S A. 2010; 107:2646–2651. [PubMed: 20133795]

Krasny L, Vacik T, Fucik V, Jonak J. Cloning and characterization of the str operon and elongation factor Tu expression in Bacillus stearothermophilus. Journal of Bacteriology. 2000; 182:6114–6122. [PubMed: 11029432]

Leadbetter JR. Cultivation of recalcitrant microbes: cells are alive, well and revealing their secrets in the 21st century laboratory. Curr Opin Microbiol. 2003; 6:274–281. [PubMed: 12831904]

Li JWH, Vederas JC. Drug discovery and natural products: End of an era or an endless frontier? Science. 2009; 325:161–165. [PubMed: 19589993]

Newman DJ, Cragg GM. Natural products as sources of new drugs over the 30 years from 1981 to 2010. Journal of Natural Products. 2012; 75:311–335. [PubMed: 22316239]

Nieselt K, Battke F, Herbig A, Bruheim P, Wentzel A, Jakobsen OM, Sletta H, Alam MT, Merlo ME, Moore J, Omara WA, Morrissey ER, Juarez-Hermosillo MA, Rodriguez-Garcia A, Nentwich M, Thomas L, Iqbal M, Legaie R, Gaze WH, Challis GL, Jansen RC, Dijkhuizen L, Rand DA, Wild DL, Bonin M, Reuther J, Wohlleben W, Smith MC, Burroughs NJ, Martin JF, Hodgson DA, Takano E, Breitling R, Ellingsen TE, Wellington EM. The dynamic architecture of the metabolic switch in Streptomyces coelicolor. BMC Genomics. 2010; 11:10. [PubMed: 20053288]

Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, Yamashita A, Hattori M, Horinouchi S. Genome sequence of the streptomycin-producing microorganism Streptomyces griseus IFO 13350. J Bacteriol. 2008; 190:4050–4060. [PubMed: 18375553]

Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, Takahashi Y, Horikawa H, Nakazawa H, Osonoe T, Kikuchi H, Shiba T, Sakaki Y, Hattori M. Genome sequence of an industrial microorganism Streptomyces avermitilis: Deducing the ability of producing secondary metabolites. Proceedings of the National Academy of Sciences of the United States of America. 2001; 98:12215–12220. [PubMed: 11572948]

Rodriguez E, Hu Z, Ou S, Volchegursky Y, Hutchinson CR, McDaniel R. Rapid engineering of polyketide overproduction by gene transfer to industrially optimized strains. J Ind Microbiol Biotechnol. 2003; 30:480–488. [PubMed: 12698320]

Schmitt-John T, Engels JW. Promoter constructions for efficient secretion expression in Streptomyces lividans. Appl Microbiol Biotechnol. 1992; 36:493–498. [PubMed: 1368202]

Shao Z, Luo Y, Zhao H. Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. Mol Biosyst. 2011; 7:1056–1059. [PubMed: 21327279]

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; 2:662–669. [PubMed: 23968564]

Shao Z, Zhao H, Zhao H. DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. Nucleic Acids Res. 2009; 37:e16. [PubMed: 19074487]

Singh S, Kate BN, Banerjee UC. Bioactive compounds from cyanobacteria and microalgae: an overview. Crit Rev Biotechnol. 2005; 25:73–95. [PubMed: 16294828]

Suarez CE, Norimine J, Lacy P, McElwain TF. Characterization and gene expression of Babesia bovis elongation factor-1alpha. Int J Parasitol. 2006; 36:965–973. [PubMed: 16677650]

Tobias Kieser, MJB.; Mark J, Buttner; Keith F, Chater; Hopwood, DA. Practical Streptomyces genetics. John Innes Foundation; Norwich, United Kingdom: 2000.

Tobias Kieser, MJB.; Mark J, Buttner; Keith F, Chater; David A, Hopwood. Practical*Streptomyces* genetics. John Innes Foundation; Norwich, United Kingdom: 2000.

Traitcheva N. Molecular analysis of the neoaureothin biosynthesis gene cluster from Streptomyces orinoci HKI 0260: a model system for the evolution of bacterial polyketide synthases. Leibniz-Institute for Natural Product Research and Infection Biology (HKI), Jena. 2006

van Wezel GP, McKenzie NL, Nodwell JR. Chapter 5. Applying the genetics of secondary metabolism in model actinomycetes to the discovery of new antibiotics. Methods Enzymol. 2009; 458:117–141. [PubMed: 19374981]

Wagner N, Osswald C, Biener R, Schwartz D. Comparative analysis of transcriptional activities of heterologous promoters in the rare actinomycete Actinoplanes friuliensis. Journal of Biotechnology. 2009; 142:200–204. [PubMed: 19464328]

Wawrzyn GT, Bloch SE, Schmidt-Dannert C. Discovery and characterization of terpenoid biosynthetic pathways of fungi. Methods Enzymol. 2012; 515:83–105. [PubMed: 22999171]

Wendt-Pienkowski E, Huang Y, Zhang J, Li B, Jiang H, Kwon H, Hutchinson CR, Shen B. Cloning, sequencing, analysis, and heterologous expression of the fredericamycin biosynthetic gene cluster from Streptomyces griseus. J Am Chem Soc. 2005; 127:16442–16452. [PubMed: 16305230]

Wilkinson CJ, Hughes-Thomas ZA, Martin CJ, Bohm I, Mironenko T, Deacon M, Wheatcroft M, Wirtz G, Staunton J, Leadlay PF. Increasing the efficiency of heterologous promoters in actinomycetes. Journal of Molecular Microbiology and Biotechnology. 2002; 4:417–426. [PubMed: 12125822]

Wingler LM, Cornish VW. Reiterative Recombination for the in vivo assembly of libraries of multigene pathways. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:15135–15140. [PubMed: 21876185]

Woodyer RD, Shao Z, Thomas PM, Kelleher NL, Blodgett JA, Metcalf WW, van der Donk WA, Zhao H. Heterologous production of fosfomycin and identification of the minimal biosynthetic gene cluster. Chem Biol. 2006; 13:1171–1182. [PubMed: 17113999]

Zerikly M, Challis GL. Strategies for the discovery of new natural products by genome mining. Chembiochem. 2009; 10:625–633. [PubMed: 19165837]

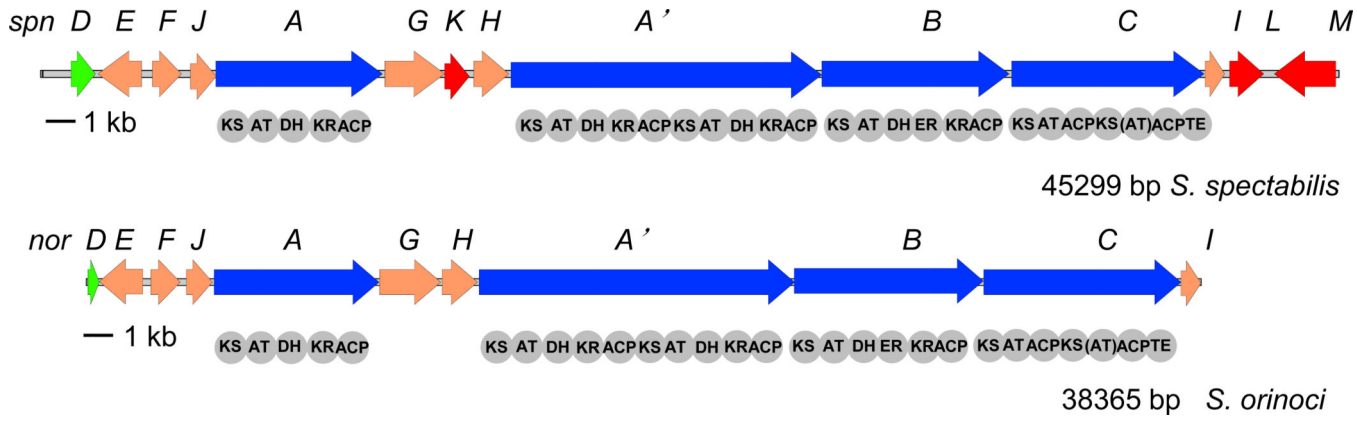


Figure 1.
The spectinabilin biosynthetic gene clusters from *S. spectabilis* and *S. orinoci*, respectively.

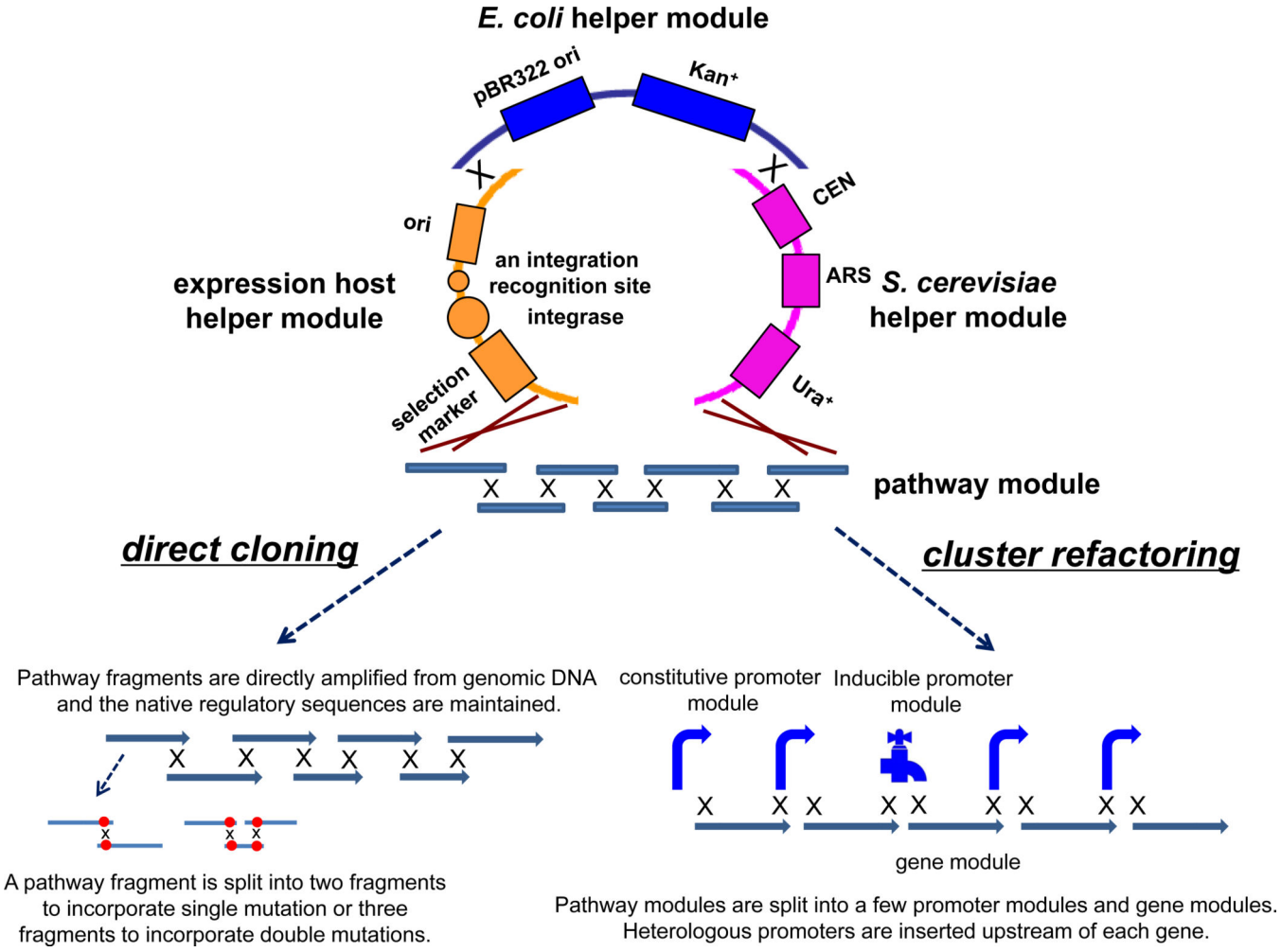


Figure 2.

The two strategies used to study natural product biosynthesis. Strategies should be selected based on the knowledge whether the native promoters could be recognized in the heterologous host. Both strategies will rely on the DNA assembler technique to assemble pathways.

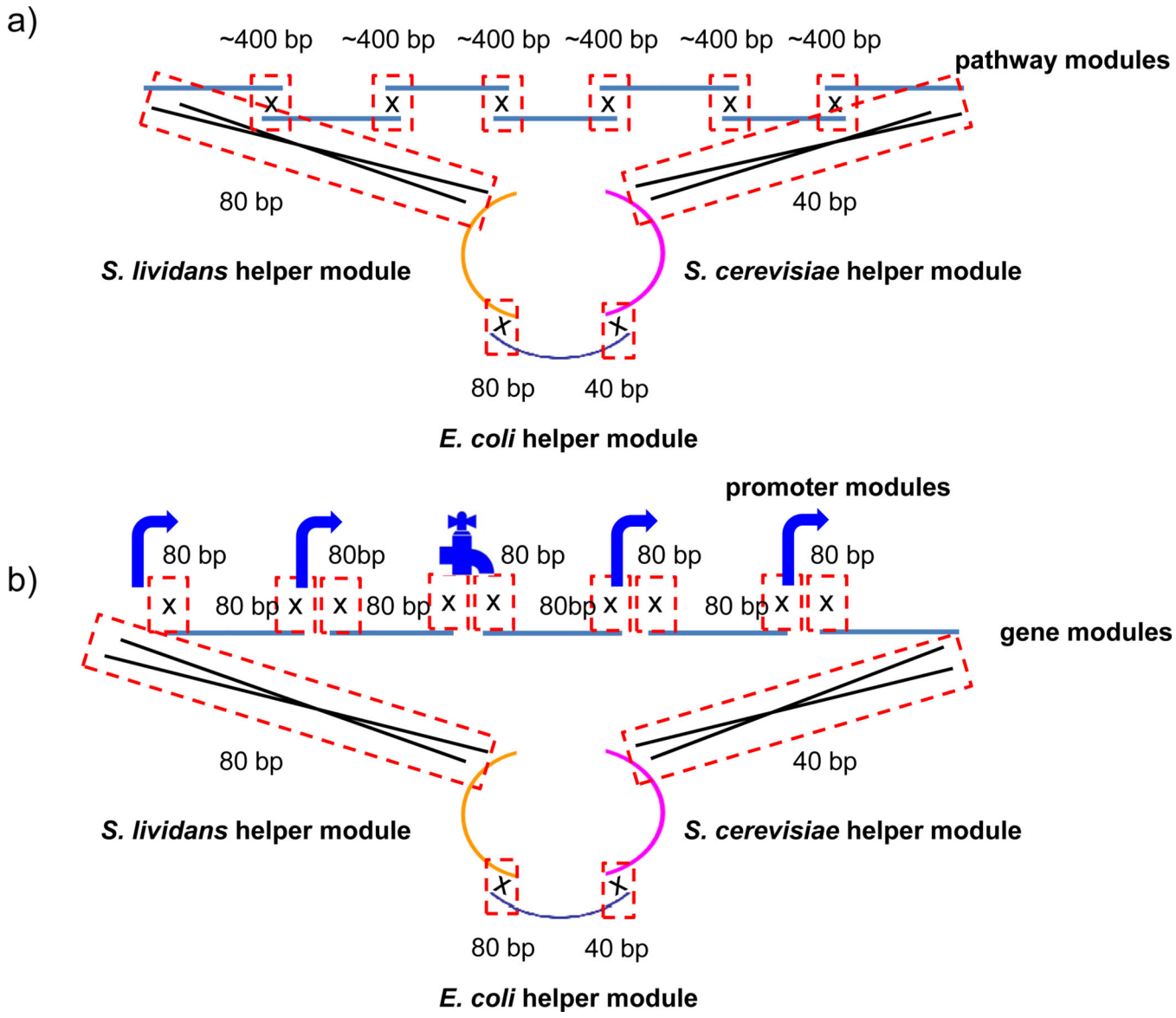


Figure 3.

The lengths of the overlaps between adjacent fragments in the direct cloning strategy (a) and the cluster refactoring strategy (b). (Shao et al. 2011) - Reproduced by permission of The Royal Society of Chemistry.

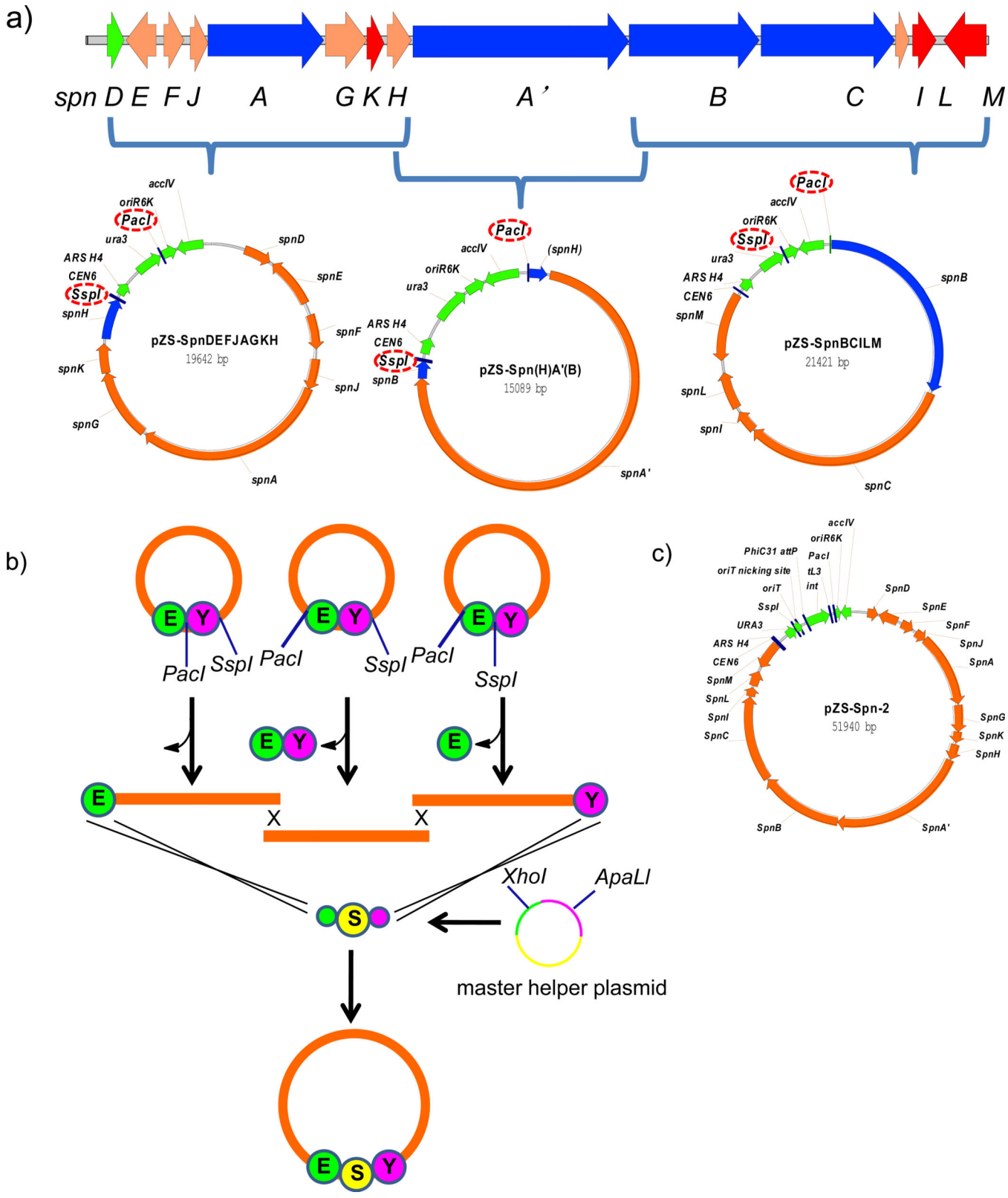
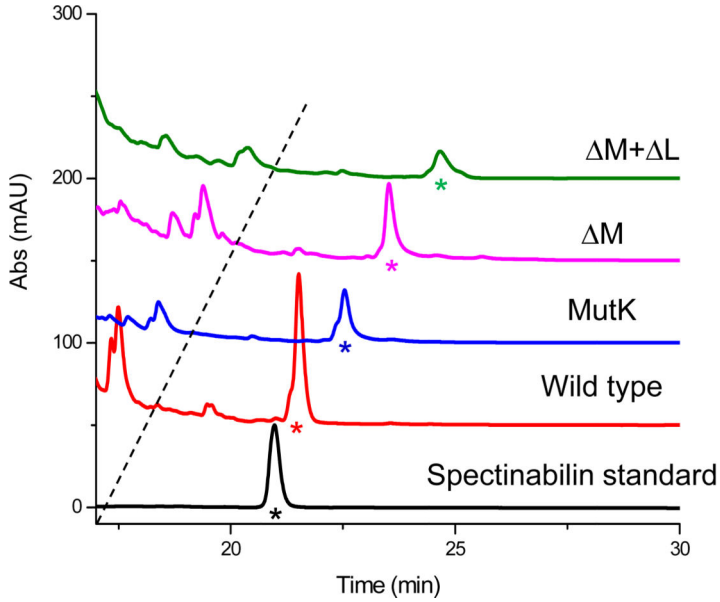
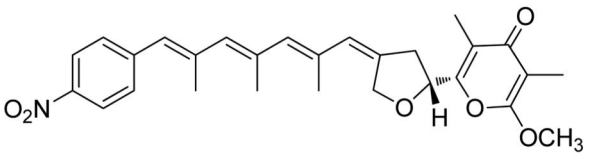


Figure 4.

The two-step strategy for assembling the spn pathway. (a) The spn cluster containing 4 PKSs are split into three intermediate plasmids. Two restriction sites, *Ssp*I and *Pac*I were engineered at the appropriate positions. (b) Restriction digestion by *Ssp*I and *Pac*I generates three intermediate fragments, which are co-transformed with a fragment obtained from restriction digestion of the master helper plasmids by *ApaL*I and *Xho*I. (c) The vector map of the assembled spn pathway. (Shao et al. 2011) - Reproduced by permission of The Royal Society of Chemistry.





20.4 min, $[M+H]^+$: 478.6
MS2: 460.2, 450.2, 400.2, 370.1, 358.1, 341.1, 302.0, 262.0, 246.1, 183.0, 167.0

Figure 5.

LC-MS analysis of the *S. lividans* clones carrying the wild type spn pathway and the mutant spn pathways. The asterisks indicate the spectinabilin peaks. MutK, the mutant with a stop codon introduced to SpnK; ΔM, the mutant with SpnM deleted; ΔM+ΔL, the mutant with SpnM and SpnL deleted. (Shao et al. 2011) - Reproduced by permission of The Royal Society of Chemistry.

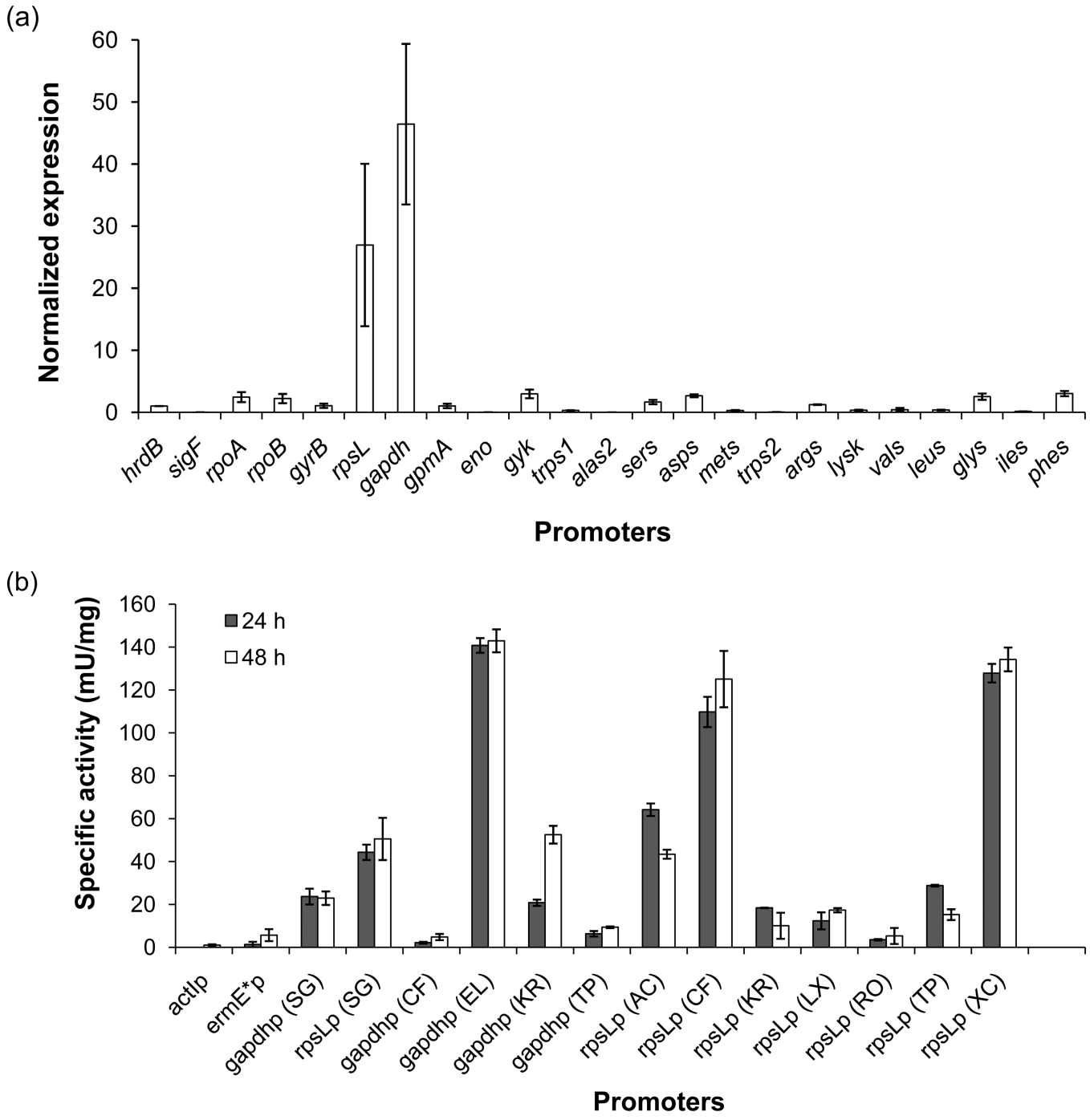


Figure 6.

Promoter screening for cluster refactoring in Streptomycetes. (a) Identification of strong constitutive promoters by real-time PCR analysis of the transcription of 23 housekeeping genes in *S. griseus*. Samples are taken at different time points. The y-axis scale represents the expression value relative to that of *hrdB*, which is set to 1. (b) Evaluation of the activities of the heterologous promoters using xylE as a reporter. actIp and ermE*p, the two commonly used promoters reported in literature (Schmitt-John and Engels, 1992; Wagner et

al., 2009; Wilkinson et al., 2002). The two letters in parentheses represent the names of individual actinomycetes. The xylE activity assay (Tobias Kieser and Hopwood, 2000) confirmed their much stronger activities than the control promoter, ermE*$_p$. Reprinted with permission from (Shao et al. 2013). Copyright (2013) American Chemical Society.

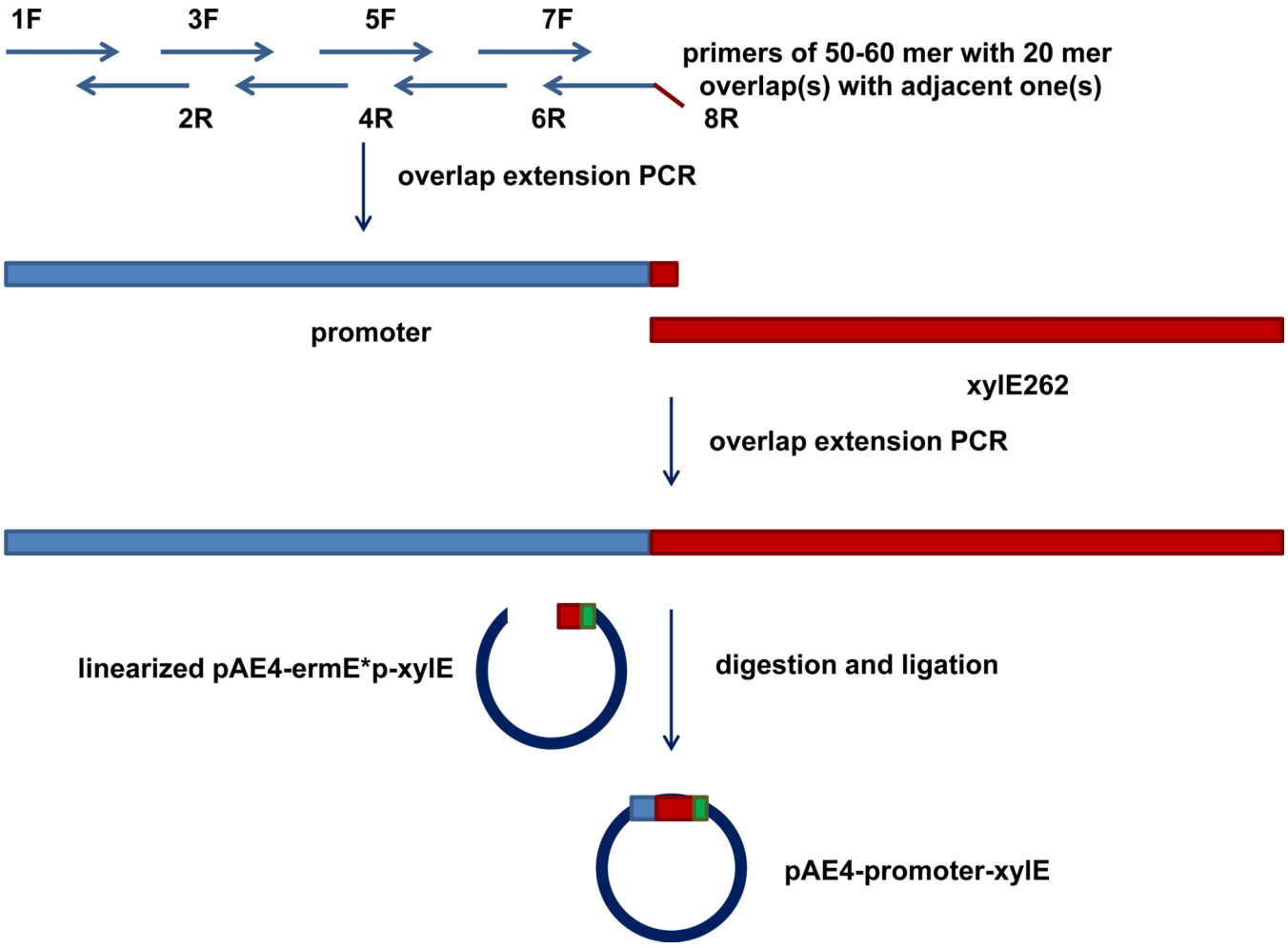
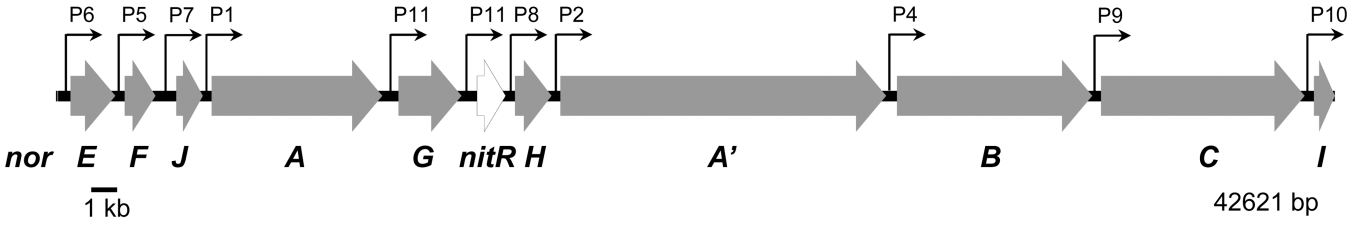


Figure 7.

Primer splicing for cloning promoters originated from actinomycetes. Reprinted with permission from (Shao et al. 2013). Copyright (2013) American Chemical Society.

(a)



P1: gapdh(KR)
P2: gapdh(SG)
P4: rpsL(TP)
P5: rpsL(SG)
P6: rpsL(CF)
P7: rpsL(AC)
P8: gapdh(EL)
P9: rpsL(KR)
P10: rpsL(XC)
P11: nitA

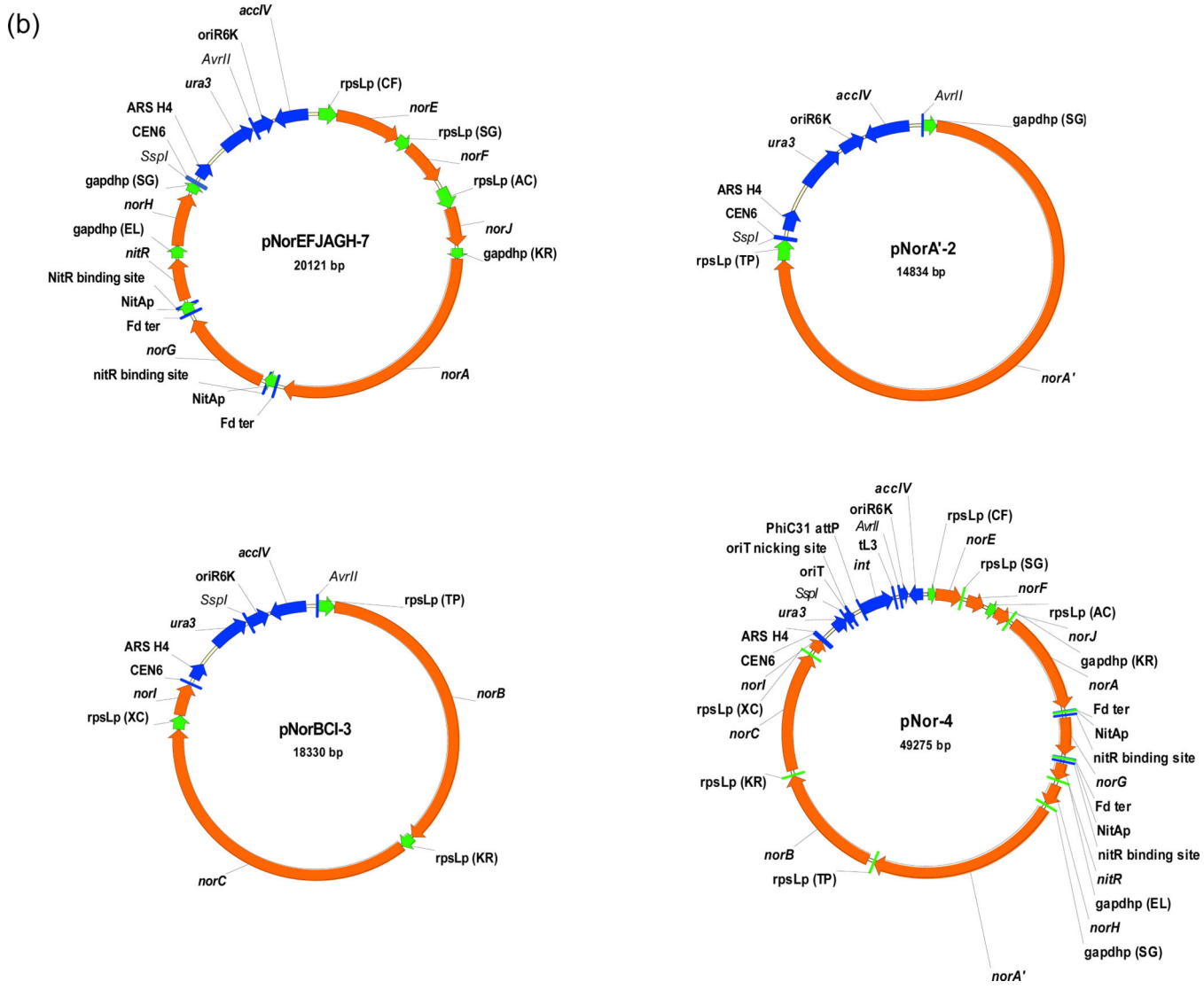


Figure 8.

(a) The refactored nor pathway. (b) The vector maps of the three intermediate plasmids and the final plasmid in the two-step assembly of the refactored nor pathway. Reprinted with permission from (Shao et al. 2013). Copyright (2013) American Chemical Society.

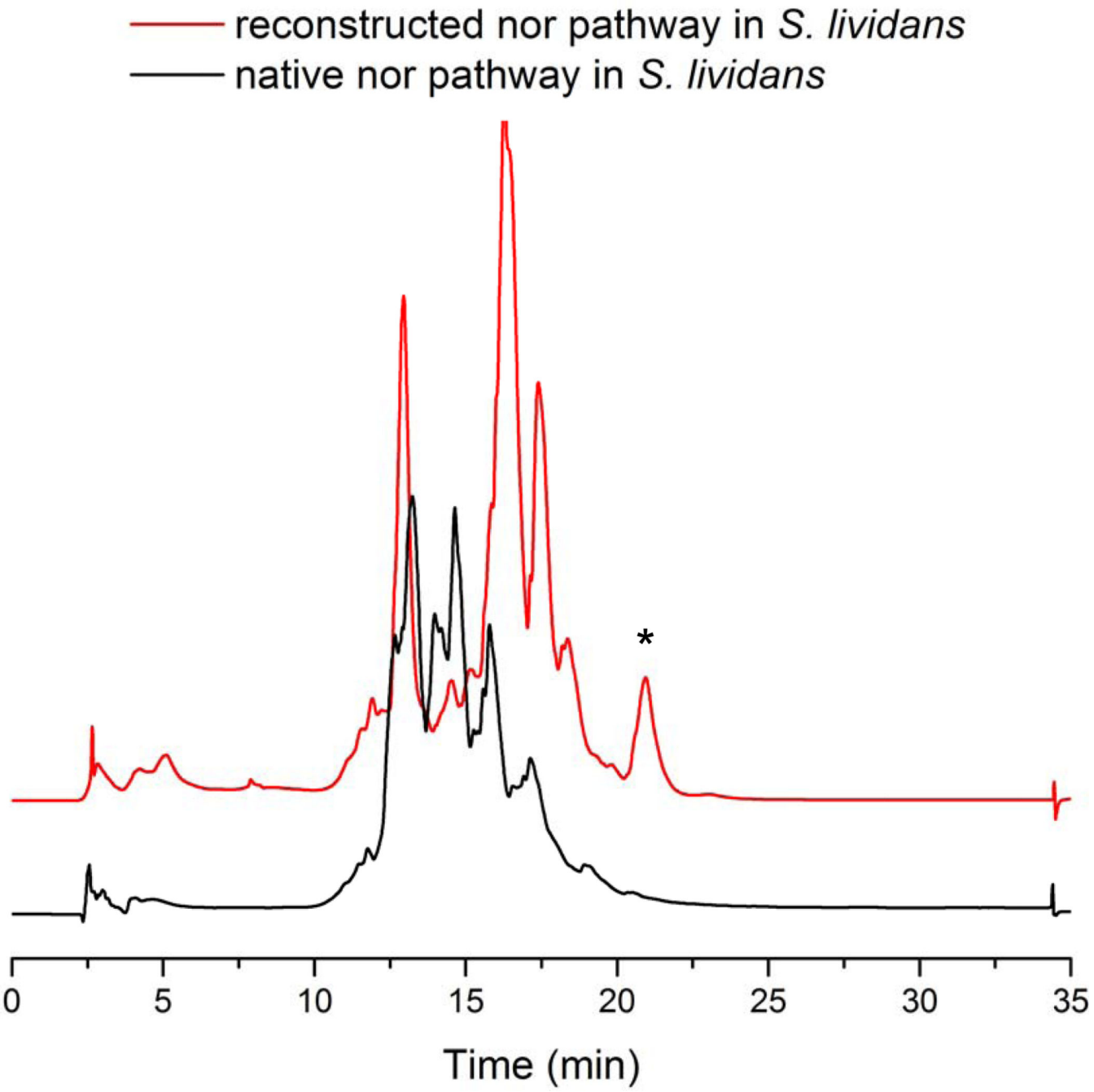


Figure 9.

HPLC analysis of the extract from the *S. lividans* strain carrying the refactored nor pathway. * indicates the spn signal peak. Reprinted with permission from (Shao et al. 2013). Copyright (2013) American Chemical Society.

Table 1

The primers used for assembling the spn cluster.

Construct	Primer name	Primer sequence (5'-3')
pZS-SpnDEFJAGKH	spn-1-for	gtagaaacagacgaagaagctagctttgcactggattgcgcccctgctcacggacgcgga
	spn-1-rev	gtgaggtcgacttcgacttc
	spn-2-for	cgcccgggacgactcgatcc
	spn-2-rev	gtccaacatcggccacaccc
	spn-3-for	ctgcgcaagggcgagtgccg
	spn-3-rev	gtcgtgacggtcgacgccct
	spn-4-for	gacgtcaccctctactcccg
	spn-4-re	tgcccggcgaacccctgtga
	spn-5-for	ccggaggtgaagcacaccgg
	spn-5-rev	tgcgcttcggcgccgcgtgaaatattcgaaaagtgccacctgggtccttttcatcacgtg
	SspI-yeast-for	aatattcgaaaagtgccacctgggtc
	PacI-yeast-rev	tgcatgtatactaaactcacttaattaa
	PacI-E. coli-for	attataagtaaatgcatgtatactaaactcacttaattaatgtcatcacgatactgtgat
	E. coli-rev	cagttaccgtgagcagatcgcccctgctcacggacgcggacgtggccgaactcatgggcc
pZS-Spn(H)A'(B)	spn-6-for	caaatacggcatcagttaccgtgagcagatcgttaattaaactcgcgccccaggcggtcg
	spn-6-rev	ccgtggagtccctcgccgta
	spn-7-for	ttggaggcgtggccgccacc
	spn-7-rev	cggagttctacgaccggatg
	spn-8-for	ggctcgccggagaccctcct
	spn-8-rev	gctcctggagaccgcctgggaatattcgaaaagtgccacctgggtccttttcatcacgtg
	SspI-yeast-for	aatattcgaaaagtgccacctgggtc
	yeast-rev	tgcatgtatactaaactcac
	E. coli-for	aaaactgtattataagtaaatgcatgtatactaaactcactgtcatcacgatactgtgat
	PacI-E. coli-rev	cagttaccgtgagcagatcgttaattaaactcgcgccccaggcggtcgaggaggtgctgc
pZS-SpnBCILM	spn-9-for	caaatacggcatcagttaccgtgagcagatcgttaattaaatgcccgctggtgacgagag
	spn-9-rev	tggctcgcggaccacgtcat
	spn-10-for	ccgcccaccctcccgatcgt
	spn-10-rev	aggtcgcggaccgccacgag
	spn-11-for	ttcgccgaggccgcgtccgt
	spn-11-rev	ggcaccgcgctcggcgaccc
	spn-12-for	gctggtggccctgcacctgg
	spn-12-rev	cacctggcaggagccgctgc
	spn-13-for	ggccctcgtcttccagcacc
	spn-13-rev	ccggccgcctcgggcagcgccgaaaagtgccacctgggtccttttcatcacgtgctataa
	yeast-for	cgaaaagtgccacctgggtc
	PacI-yeast-rev	tgcatgtatactaaactcacaatatt
	SspI-E. coli-for	gtattataagtaaatgcatgtatactaaactcacaatatttgtcatcacgatactgtgat
	PacI-E. coli-rev	cagttaccgtgagcagatcgttaattaaatgcccgctggtgacgagaggttcgtcgaggc

Construct	Primer name	Primer sequence (5'-3')
pZS-SpnMutK	spnK-stop199-rev	tgcggcgcctcgccgacgac*tga*cgcacgctcttcggcacggacccggacaccgcgaccg
	spnK-stop199-for	ggccgcgtccgcgggctcggcctgcatctgcggcgcctcgccgacgac*tga*cgcacgctcttcggcac
pZS-SpnBCIM	spnM-rev	agggatggcaggtcgcgcgtcgaaaagtgccacctgggtccttttcatcacgtgctataa
pZS-SpnBCI	spnI-rev	gtacacgctccggagtcgcccgaaaagtgccacctgggtccttttcatcacgtgctataa

Table 2

The list of the 23 housekeeping genes from *S. griseus* and the locations for the corresponding potential promoters. Reprinted with permission from (Shao et al. 2013). Copyright (2013) American Chemical Society.

Function category	Gene	Protein	Strand	Promoter location	Promoter length (nt)
gene transcription	hrdB	RNA polymerase sigma factor	−	2007417–2007753	337
	sigF	RNA polymerase sigma factor F	+	4155085–4155283	199
	rpoA	RNA polymerase subunit α	−	3319939–3320075	137
	rpoB	RNA polymerase subunit β	−	3374401–3374993	593
DNA replication	gyrB	DNA gyrase subunit B	+	4329454–4329912	459
translational elongation	rpsL	30s ribosomal protein S12	−	3350362–3350672	311
glycolysis	gapdh	glyceraldehydes-3-phosphate dehydrogenase	+	6550448–6550686	239
	gpmA	phsphoglycero- mutase	−	4689252–4689458	207
	eno	enolase	+	5190108–5190429	322
	gyK	pyruvate kinase	+	6481483–6481601	120
amino-acyl tRNA synthetase	trps1	tryptophanyl-tRNA synthetase	−	3194954–3195145	192
	alas2	alanyl-RNA synthetase	+	3590598–3590924	327
	sers	seryl-tRNA synthetase	+	4242374–4243105	732
	asps	aspartyl-tRNA synthetase	+	4422111–4422277	167
	mets	methionyl-rRNA synthetase	−	4426808–4426991	184
	trps2	trptophanyl-tRNA synthetase	+	4445297–4445644	348
	args	arginyl-tRNA synthetase	−	4885112–4885267	156
	lysk	lysyl-tRNA synthetase	+	4885112–4885267	156
	vals	valyl-tRNA synthetase	+	5804700–5804837	138
	leus	leucyl-tRNA synthetase	+	5866038–5866611	574
	glys	glycyl-tRNA synthetase	+	5921526–5921701	176
	iles	isoleucyl-tRNA synthetase	−	6385981–6386544	564
	phes	phenylalanyl-tRNA synthetase α	+	6913456–6913570	115

Table 3

The primers used for cloning $gapdh_p$ (SG) and $rpsL_p$ (SG) upstream of *xylE*. Restriction sites are underlined.

Construct	Primer name	Primer sequence (5'-3')
pAE4-$gapdh_p$ (SG)-xylE	AflII-$gapdh_p$ (SG)-for	atgcatgccttaagctcggcggcgagcgggggac
	$gapdh_p$ (SG)-rev	cgcattacacctttgttcatgaaccgatctcctcgttagg
	$gapdh_p$ (SG)-xylE-for	cctaacgaggagatcggttcatgaacaaaggtgtaatgcg
	NdeI-xylE-rev	ggcacagccatatgccatcagatcccgctccagt
pAE4-$rpsL_p$ (SG)-xylE	AflII-$rpsL_p$ (SG)-for	atgcatgccttaagaggggcgcgcggccccggcc
	$rpsL_p$ (SG)-rev	cgcattacacctttgttcattacttctccggtttctgtgt
	$rpsL_p$ (SG)-xylE-for	acacagaaaccggagaagtaatgaacaaaggtgtaatgcg
	NdeI-xylE-rev	ggcacagccatatgccatcagatcccgctccagt

Table 4

The sequences of the heterologous promoters which are active in *S. lividans*. Reprinted with permission from (Shao et al. 2013). Copyright (2013) American Chemical Society.

Promoter	Origin	Length (nt)	Sequence
$actI_p$	*Streptomyces coelicolor*	1505	aagcttagatctcgagtttaaacagctcggattcacggtcgcggtgggcgtcctgctcgacaccatgatcgtgcgatcggtcctgg tcaccgcgctcactctcgatgtcggccggtggatgtggtggccgcacccgctcgcccggcgcgaggacccttccgaggaccca gccgtatcaggaatgccagattctattgattcggaagcctcgaccactgcctctcggtaaaatccagcaaaaattaatcagtgca gctcgctgcactgattaatttttgatcaataggagatcgcttgtgacggcaagcacattgaaatctgttgagtaggcctgttattgtcg cccccaggagacggagaatctcgacgggggcgcagatgagattcaacttattgggacgtgtccatgtaatcaccgatgcggga tgtgtaattccgcttaaatcctcgaaggcgacccagctcctggtgctgctgctcctcaggcggcacgaggtggtgggatcggggg tgctcatcgaggagttgtgggcggaccacccgccccgcagcgccatgacgacgctgcagacgtacgtgtaccacacccgccg gctgctgggggagcaccgggtgacgagcgacgaccgggaattggtcctgacccagccgccggctacttcgccctgatcgac gaggacgaactcgacgtcgcggtcgccgagcgtctgatccgcaccggcggccggctgctcgaggagaaccggctcgagga ggcgctcgcctcgttggacgcgggactggatctctggcgaggcccggcgctgtccaccgtaccgtgcggccgggtgctcgaaa gcaatatcgcgcacctggaagagctgcggctttttggaatgcagctccgtatcgacgcgaattggcggctgggcagaatagggc cgatgattccggaactccggtccctggtaatttcgcatccgctgaacgagaccctgcacgccaaactgatgggcgcgctctgtca gatgggcaggcgcgccgaggcgctggaatcgtatcggaatctccggcggatactgtccgacgaactgggggtggatccgacg ccggaaatccagcgtatgcacatggaaattctcaacggtgagaaggtgctcgtgtagcaccggtccgtgaacgcggtggagcc cctcgcagcctgcgaggggctccaccgcgttcacggaccggccggcgtcatttaaatctagaggttcgaatgcatatgttcccct ccctgcctcgtggtccctcacgcgctcagctttgggcgcccggctcgagcggcggtcgaagggagatggggtgccgctggacg cggcgccggtggatccggcatcgaggggtcccgtatcggccttcgagcctccttcgagccacggggccgacgatgacgacga ccaccggacgaacgcatcgattaattaaggatccggcgcgccccgggtaccgagctcgaattctcgcgagagaggcgtatc c	
$ermE^*_p$	*Saccharopolyspora erythraea*	282	ggtaccagcccgacccgagcacgcgccggcacgcctggtcgatgtcggaccggagttcgaggtacgcggcttgcaggtcca ggaaggggacgtccatgcgagtgtccgttcgagtggcggcttgcgcccgatgctagtcgcggttgatcggcgatcgcaggtgca cgcggtcgatcttgacggctggcgagaggtgcggggaggatctgaccgacgcggtccacacgtggcaccgcgatgctgttgtg ggctggacaatcgtgccggttggtaggacaccac
$gadph_p$ (SG)	*Streptomyces griseus*	239	ctcggcggcgagcgggggacgttggtcccgcacccgccgggacccgggtcccggcgggcgaaacggaggtccgtgctcct actggcgagtacggatctccggccgtatccggtggcgtgcggtggccctctcgatgtcaccccggggcctcggagaggtagg gtcggaagcggtcggggacatccctatacaactcgccggcgtcgaaaaccggcgtacctaacgaggagatcggttc
$rpsL_p$ (SG)	*Streptomyces griseus*	311	aggggcgcgcggccccggccgccggccgtcccgtccgggggcccgtggccggggcggtcccggcgtgtcgcggcggaca gcatttgttttgacccagctccgtgaggtaggtacgctcaagccttgtgcctggggtgtgcctgggctcgggtgcgtgtcctcaacc gcatcgcgagtccgtcagtagccaccgcaatctgcgcccttcctgccttcggggcgggagtccgcagtattcgacacacccgac cgcgtgggtcggcgatgttccaggttagtttcacgaacggcacacagaaaccggagaagta
$gapdh_p$ (CF)	*Cellulomonas flavigena*	252	caccggggcgcccgtcgtcggcgcgcccgcgtggtccagcatgcgaacgcaggtccgggccggaccttcgtcttggccgacg ccctcgccgcggggtataggctcgggtcatccgatcgcaacgacgacgcaacgcgctgggcacatcccggcggtggctggtc ggcggcggccggaacgcacagcggcgtgcgtgacccgaacaccaactgtgtgcgccggcatgggatcggcgcgtgccgag gagggcaa
$gapdh_p$ (EL)	*Eggerthella lenta*	285	gctgctccttcggtcggacgtgcgtctacgggcaccttaccgcagccgtcggctgtgcgacacggacggatcgggcgaactgg ccgatgctgggagaagcgcgctgctgtacggcgcgcaccgggtgcggagcccctcggcgagcggtgtgaaacttctgtgaat ggcctgttcggttgctttttttatacggctgccagataaggcttgcagcatctgggcggctaccgctatgatcggggcgttcctgcaat tcttagtgcgagtatctgaaaggggatacgc
$gapdh_p$ (KR)	*Kocuria rhizophila*	235	Cacgggccggtgaaccccgcatgaatgttggttcatgtggggtttgcgccagctgtgtcgcagttgccccccgccactgttcccg gggcatgaatcccgctagaatgaacacggaagagcaggtggaccccgggtcgtcaccgcacttcgggtgtgggatttatgccg attggttttgatcccacatccggacgggccggagtgcaccgcacacccactcgcagaggagattgca
$gapdh_p$ (TP)	*Tsukamurella paurometabola*	162	Tccgttcccggcggggagcgggacacccggggagttcgcccggaaatgtccgaagctaggcgaagaaagtgtgacgagca cgaagtccctgctccctgccggtgcgaatgtgccacgcgatagggtgtggtgacaggcacacccagtcgaaggagcagttaa
$rpsL_p$ (AC)	*Acidothermus cellulolyticus*	490	gcgcggtctcgccggccggcggcggtgcccggttgttcactccgacggctccgggagtcgccggagccgtcgccgagaaga gggccggtgcgcttcgtcattggcgcaggcgttccccgtgccgtcgagcggtgggttgctccgtgacgtcgcggtcccggtggc tggcgtagccgcgaactacggcagcgactgaccgctgctgcggccggctactcggcgcgccggcggagtccctgcgctgaa acgggacccggttttgacctgccggcacgggctcggtaacctagtcgctcgtgcccggcttcccgggcgccttcgctgttctgcgc gcgttgcttggtttgcgcggtgctcggcgtgtggccgccggcaggcggcctccggccgagcccgacacgcccgaccgcgtgga gcggcgatagggtgcagggagcaggacatcggacagaccgagcgatcacgtttgaggatccacgaagggcgcgt
$rpsL_p$ (CF)	*Cellulomonas flavigena*	419	cccgccgcgggcgctggaggctcgggcgggcccgggccggaggcggccgcgaccacgacgcccgcgggacgtgacga gcggcacgactcgacgactccgggctcctttgacgctgtccgtcgccgggtagcgtaggacaccgtgcccgcgccgtcggg ccctcgcgcgtgcactcggtcgaccgctccctgccggagtgggtgcgggtgcacggggtggctccccacctcctctcggatcgg tcctcgcggactgccgccgtgcggaggaccggggcgacacgcccgggcgcgggggtcggtgcgggactccagacctccgg ggtagtcgtgcgacgggcgacgatccgggccgagccggccgtcctgggtgacgggtgccggtcagaccagagaacaccga cagacggagacgta

Promoter	Origin	Length (nt)	Sequence
$rpsL_p$ (KR)	*Kocuria rhizophila*	286	tcacgcgtggctgggtgcgggctcggtggtctggggcgtcggtttcccctgagctgccgtggctgatacgatgggccagaattgct cgattccgggcagtttgcaatgagatggcaaatcggactccgtcctgtggtacgcaccatgtacctagatgccatctttcgtcatgc tcaggtgcggtttccgaggcgttcgccccggcacgcaccgggcaccccgggatcaacgttccgtcatcggcggggccggcac acaccgcgcacccgccccaacaggaggcact
$rpsL_p$ (LX)	*Leifsonia xyli*	359	gggcaggggcggagcggaccgggagagccggcgggacaccacgctctcttcgcaacctcagggggaagcgccggtagcat gcccgcgatcatttgaccagctttcgcgtggacgatagtatttatcgggtgtgcgctctgtcgcgcgcttgtcctatggtcactgaccgg ctcggtccgggactgggaggagcgggccggcacgtactcatccgatgggctgaggagaccgcacgggaacccgcccccacg gcatatcggcagctcgccgatcaccccacgaccagaggcgccgcgaggcgcggctgggccgtgggaaagcaacaactgtca ccgtgcagtcctaaaaggagaagttca
$rpsL_p$ (RO)	*Rhodococcus opacus*	362	tgtgcccttgcccgtaccgacagctccgagcagttcccgacatcgcaggtcagcctttaccattgtccgcccactttgaccttccggc gcgccgccgggtatcgttgtgggtcgtgtccggcatgcccgggccgatctgtgtgcctagcgaggtattgccgtgcgcgcgttcgtg cagatttctgcatggaacgagcatccggctgcctgtgtgggggtatgcgacacacccgacctcgggtgcagctcgaggtgcgggt gaggggcagcagctctcttgccccaactgctagatccctaatttcagaacagccaaggttgcttgtccagagcagtctgttcaacac gaagaaagccggtaa
$rpsL_p$ (TP)	*Tsukamurella paurometabola*	363	accgggtccgcgatcggcggaggcgaacacttttgaccactatgagttcatccaggtaagcttggtcgccgtgcctggtgaatgcc aggacgatcgctcgtgcccatactgcaggccggacctccgggaccacgtacggacacgcgacacgcccgacctcggggtacg tgcgaggcgggtagctacttcccggaaacgggactgaccagcagagacggcgaaagccggaactgccggtgggggccgcgct gcggaagtcgcaccctcgatgagggacggcttgcccggagctacaaccgcagcagtagagccaccggccggacggccgatg gcagccgaaacgaagtaaggaacctgcgcttct
$rpsL_p$ (XC)	*Xylanimonas cellulosilytica*	302	gccctgcaggcggaagtcaggtagacacgacttccgctagtccttgcaaggtctgctgacgtgaggcggggcggtcgtttttgacc gccccgccttcgtcatgtaggctcgctcgctgtgcctggcgtgtcttcagacgcccaggtcccggtgccgtgaggcccgggccatcg agccggtggtacgtggctgcggtcccctttgtgagggctgcgcgccgtgtgctgtccggcgcgcacagccttgaatccacccgcgg gggccggccggtctccgtgagctcgagaagacgacggagacgta

Table 5

The primers used for refactoring the nor cluster. Reprinted with permission from (Shao et al. 2013). Copyright (2013) American Chemical Society.

Construct	Primer name	Primer sequence (5'-3')
pNorEFJAGH-7	N1-rpsLp (CF)-for	tggtactgcaaatacggcatcagttaccgtgagcagatcg cccgccgcgggcgctggagg
	rpsLp (CF)-rev	cgcggtcgaggaagtcggccatggtgagcggaacgagcattacgtctccgtctgtcggtg
	NorE-for	gtgccggtcagaccagagaacaccgacagacggagacgtaatgctcgttccgctcaccat
	NorE-rev	cggggccgcgcgcccctttagttgatccggcgctcctggc
	rpsLp (SG)-for	gccaggagcgccggatcaactaaaggggcgcgcggccccg
	rpsLp (SG)-rev	tccgcctgcgcttctcgcattacttctccggtttctgtgt
	NorF-for	acacagaaaccggagaagtaatgcgagaagcgcaggcgga
	NorF-rev	gtgaacaaccgggcaccgccgccggccggcgagaccgcgcggtgaggaaccgctgccctt
	rpsLp (AC)-for	cctcgccgcggcctcgccggaagggcagcggttcctcaccgcgcggtctcgccggccggc
	rpsLp (AC)-rev	tgtccacgaccgtctccaccgccctggtggcacgcagcacacgcgcccttcgtggatcct
	NorJ-for	agaccgagcgatcacgtttgaggatccacgaagggcgcgtgtgctgcgtgccaccagggc
	NorJ-rev	gcggggttcaccggcccgtgggttgcttccttcctggcag
	gapdhp (KR)-for	ctgccaggaaggaagcaacccacgggccggtgaaccccgc
	gapdhp (KR)-rev	cgtatcccgttcaactccattgcaatctcctctgcgagtg
	NorA-for-1	cactcgcagaggagattgcaatggagttgaacgggatacg
	NorA-rev-1	ccagcacccggtggccatgg
	NorA-for-2	gagtacggcgcccgcatccg
	NorA-rev-2	tccaaaaggagcctttaatttgtaccgcctgcacgaagtgcgggacgtcctcactcgccg
	NitAp-for	gactttatcgaccaccagctcggcgagtgaggacgtcccgcacttcgtgcaggcggtaca
	NitAp-rev	aaattatttctagagtcgacctgcaggcatgcaa
	NorG-for	ggtcgactctagaaataattttgtttaactttaagaaggagatataccatgagcgtgctgctgatcga
	NorG-rev	tcaaaggccgcttttgcggggaacgcgtagatctgaattctcagtgggacttggcgcctt
	NitR-for	cggtcccagccgggtccaagaaggcgccaagtcccactgagaattcagatctacgcgttc
	NitR-rev	gtaaggtgcccgtagacgcacgtccgaccgaaggagcagcctctagtccgggctcttcct
	gapdhp (EL)-for	atcaccgacggaggtttcgtaggaagagcccggactagaggctgctccttcggtcggacg
	gapdhp (EL)-rev	ggtcctgccagggggattcggtggtcggggcggtagacatgcgtatccccctttcagatac
	NorH-for	tcctgcaattcttagtgcgagtatctgaaaggggatacgcatgtctaccgccccgaccac
	NorH-rev	gtcccccgctcgccgccgaggtctaatttcgctgcccttc
	gapdhp (SG)-for	gaagggcagcgaaattagacctcggcggcgagcgggggac
	SspI-gapdhp (SG)-rev	cacgtgatgaaaaggacccaggtggcacttttcgaatattgaaccgatctcctcgttagg
	N1-SspI-yeast-for	aatatt cgaaaagtgccacctgggtc
	N1-AvrII-yeast-rev	cctagggtgagtttagtatacatgca
	N1-AvrII-E. coli-for	attataagtaaatgcatgtatactaaactcaccctaggtgtcatcacgatactgtgat
	N1-E. coli-rev	cggcccggggcccgcccgagcctccagcgcccgcggcgggcgatctgctcacggtaactg
pNorA'-2	N2-AvrII-gapdhp (SG)-for	Tgcaaatacggcatcagttaccgtgagcagatcgcctaggctcggcggcgagcgggggac
	gapdhp (SG)-rev	gtcttgtcgtccttggtcatgaaccgatctcctcgttagg
	NorA'-for-1	cctaacgaggagatcggttcatgaccaaggacgacaaga

Construct	Primer name	Primer sequence (5'-3')
	NorA'-rev-1	ccagcagagccgggtggagg
	NorA'-for-2	gccgaactgacccttcaggc
	NorA'-rev-2	cctcgatcggatcgcccagg
	NorA'-for-3	gtgcgcgccggggactgcga
	NorA'-rev-3	gtggtcaaaagtgttcgcctccgccgatcgcggacccggttcaccaaccggcatcgtggt
	rpsLp (TP)-for	agcacacgccgagacggaggaccacgatgccggttggtga accgggtccgcgatcggcgg
	SspI-rpsLp (TP)-rev	cacgtgatgaaaaggacccaggtggcacttttcgaatattagaagcgcaggttccttact
	N1-SspI-yeast-for	aatattcgaaaagtgccacctgggtc
	yeast-rev	gtgagtttagtatacatgca
	E. coli-for	aaaactgtattataagtaaatgcatgtatactaaactcactgtcatcacgatactgtgat
	N2-AvrII-E. coli-rev	ggtgcgggaccaacgtcccccgctcgccgccgagcctaggcgatctgctcacggtaactg
pNorBCI-3	N3-rpsL (TP)-for	tgcaaatacggcatcagttaccgtgagcagatcg cctaggaccgggtccgcgatcggcgg
	rpsL (TP)-rev	cccgcagtgcctcgacgagcttgtcgtcaccaaccggcatagaagcgcaggttccttact
	NorB-for-1	gccgatggcagccgaaacgaagtaaggaacctgcgcttctatgccggttggtgacgacaa
	NorB-rev-1	tgaggaaggtgacggggacg
	NorB-for-2	ggaaggttgctggtgcccag
	NorB-rev-2	gacgccccagaccaccgagcccgcacccagccacgcgtgacggcggtcaggagatgccga
	rpsLp (KR)-for	cgacttcatcagcaatgaactcggcatctcctgaccgccgtcacgcgtggctgggtgcgg
	rpsLp (KR)-rev	cgtccgtcaccatctgagaaacaccccgtccagaagtcatagtgcctcctgttggggcgg
	NorC-for-1	Ggccggcacacaccgcgcacccgccccaacaggaggcactatgacttctggacggggtgt
	NorC-rev-1	tggacatgggcttcgccgac
	NorC-for-2	tgcgagccatccgggtcgcg
	NorC-rev-2	ctagcggaagtcgtgtctacctgacttccgcctgcagggctcaatgctccttttttgaa
	rpsLp (XC)-for	ccaccactggctcaacggccttcaaaaaaaggagcattgagccctgcaggcggaagtcag
	rpsLp (XC)-rev	ggtcgaggggcttttcctggctggtgccgctggacgtcattacgtctccgtcgtcttctc
	NorI-for	cggccggtctccgtgagctcgagaagacgacggagacgtaatgacgtccagcggcaccag
	NorI-rev	ttatagcacgtgatgaaaaggacccaggtggcacttttcgctatgcggcgccggtccagg
	yeast-for	cgaaaagtgccacctgggtc
	SspI-yeast-rev	aatattgtgagtttagtatacatgca
	SspI-E. coli-for	gtattataagtaaatgcatgtatactaaactcacaatatttgtcatcacgatactgtgat
	N3-AvrII-E.coli-rev	aaaagtgttcgcctccgccgatcgcggacccggtcctaggcgatctgctcacggtaactg

Table 6

The anticipated time required for each major step.

Step	Estimated time	Explanation
Preparation of fragments by PCR	1 day to 1 week	For certain fragments, reaction conditions need to be fine-tuned.
Yeast transformation	2.5 days	The transformation step itself takes 6 hours and plates are incubated for 2 days for colonies to appear.
Construct verification	4–5 days	Individual yeast colonies are inoculated into fresh medium and grown for 1 to 2 days, after which, plasmids are purified and transformed into *E. coli*. *E. coli* colonies are inoculated into fresh medium and grown for overnight. Plasmids are subsequently purified and verified.
Conjugation	1 week	Verified constructs need to be transformed to *E. coli* donor cells first. The conjugation experiment takes 4–5 days for exconjugants to appear.
Recombinant clone culturing	1 week	Cells are grown for approximately 1 week and samples are taken at appropriate time.
Product detection	1–2 days	Product is extracted by ethyl acetate and subjected to HPLC analysis.
Promoter screening	4–5 weeks	Strong promoters are usually identified *via* real-time PCR analysis of the samples prepared in the selected conditions. Cloning the candidates upstream of the reporter gene and confirming their activities take about 3 weeks.