Published in final edited form as:
Methods Enzymol. 2012 ; 517: 225–239. doi:10.1016/B978-0-12-404634-4.00011-5.

Reassembly of functionally intact environmental DNA-derived biosynthetic gene clusters

Dimitris Kallifidas and **Sean F. Brady***
Laboratory of Genetically Encoded Small Molecules, Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, NY 10065

Abstract

Only a small fraction of the bacterial diversity present in natural microbial communities is regularly cultured in the laboratory. Those bacteria that remain recalcitrant to culturing cannot be examined for the production of bioactive secondary metabolites using standard pure-culture approaches. The screening of genomic DNA libraries containing DNA isolated directly from environmental samples [environmental DNA (eDNA)] provides an alternative approach for studying the biosynthetic capacities of these organisms. One drawback of this approach has been that most eDNA isolation procedures do not permit the cloning of DNA fragments of sufficient length to capture large natural product biosynthetic gene clusters in their entirety. Although the construction of eDNA libraries with inserts big enough to capture biosynthetic gene clusters larger than ~40 kb remains challenging, it is possible to access large gene clusters by reassembling them from sets of smaller overlapping fragments using transformation associated recombination (TAR) in *Saccharomyces cerevisiae*. Here we outline a method for the reassembly of large biosynthetic gene clusters from captured sets of overlapping soil eDNA cosmid clones. Natural product biosynthetic gene clusters reassembled using this approach can then be used directly for functional heterologous expression studies.

Keywords

Natural products; metagenomics; yeast recombination; *Streptomyces*; heterologous expression

1. Introduction

Culture-independent analyses of natural microbial populations, including 16S rRNA sequencing and shotgun sequencing of DNA extracted directly from environmental samples, has consistently shown that only a small fraction of bacterial species is readily cultivated under standard laboratory conditions (Rappe and Giovannoni, 2003; Torsvik *et al.*, 2002; Torsvik *et al.*, 1990). By some estimates, as much as 99% of the bacterial diversity present in many environmental samples is not readily cultured, rendering these organisms ineffectual as sources of new bioactive small molecules. The cloning and subsequent analysis of DNA extracted directly from environmental samples, which has broadly been defined as the field of metagenomics, is now being used to unravel the hidden biosynthetic capacity of natural bacterial populations (Handelsman *et al.*, 1998; Brady *et al.*, 2009; Daniel, 2005; Simon and Daniel, 2009). In bacteria, genes responsible for the biosynthesis of secondary metabolites, including genes with biosynthetic, transport, resistance and regulatory functions, are often found clustered on the bacterial chromosome (Bentley *et al.*,

*Corresponding Author: Sean F. Brady, Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, 1230 York Avenue, New York, NY 10065, Phone: 212-327-8280, Fax: 212-327-8281, sbrady@rockefeller.edu.

2002). This conserved architecture can be exploited to permit the cloning of complete secondary metabolite biosynthetic gene clusters directly from environmental samples. The analysis of eDNA libraries has now identified new biosynthetic enzymes, gene clusters and bioactive small molecules.

While the construction of cosmid-based eDNA libraries containing 30–40-kb inserts is now routine, it remains technically challenging to construct larger insert libraries that would permit the cloning of biosynthetic gene clusters that span more than 40 kb of DNA. Bacterial artificial chromosome (BAC)-based metagenomic libraries offer the possibility of capturing DNA inserts longer than 100 kb. However, because of the difficulties associated with obtaining very large DNA fragments from environmental samples they are often orders of magnitude smaller than comparable cosmid-based libraries (Daniel, 2005). An alternative strategy to capturing complete biosynthetic gene clusters on individual large insert clones is to recover the clusters on multiple overlapping clones and then reassemble the pathway from these overlapping fragments. Restriction enzyme- and bacterial recombination-based approaches can be used to overcome simple reassembling problems (Binz *et al.*, 2008) but these approaches are laborious and often not practical when faced with the need to reassemble very large gene clusters captured on multiple overlapping clones. Yeast has a more efficient DNA recombination system than most bacteria (Nagano *et al.*, 2007; Larionov *et al.*, 1994) and we have found it very useful for reassembling collections of eDNA cosmids into complete biosynthetic gene clusters.

Transformation-associated recombination (TAR) in *Saccharomyces cerevisiae* was initially developed as a way to permit the direct cloning of target DNA fragments from pools of crude genomic DNA (Mathee *et al.*, 2008; Kouprina and Larionov, 2008; Larionov *et al.*, 1996; Gibson *et al.*, 2008a; Kouprina *et al.*, 2006; Nagano *et al.*, 2007). In TAR cloning protocols, genomic DNA and a capture vector, engineered to carry small homology arms corresponding to sequences flanking the target DNA of interest, are introduced into *S. cerevisiae* by co-transformation. The capture vector arms and the homologous genomic sequences undergo recombination to yield a stable plasmid containing the targeted genomic region. In more recent years, this technique has been exploited to assemble sets of overlapping synthetic DNA fragments into larger sequences (Gibson *et al.*, 2008b; Shao *et al.*, 2009). The protocol outlined here describes the creation of large eDNA libraries from soil, the arraying of these libraries to facilitate the recovery of overlapping clones and the use of TAR to reassemble overlapping clones into full-length gene clusters that can be used in downstream heterologous expression studies (Figure 1).

2. Protocol for constructing, arraying and screening environmental libraries

2.1. Isolation of high-molecular-weight eDNA and library construction

Methods for constructing large eDNA libraries from environmental samples have been described in detail elsewhere (Brady, 2007; Rondon *et al.*, 2000; Simon and Daniel, 2010; Zhou *et al.*, 1996). Here we briefly outline a strategy for constructing very large eDNA cosmid libraries from freshly collected soil samples. In brief: 250 g of soil which has been passed through a 1/8″ (3–4 mm?) screen is mixed 1:1 (w:v) with lysis buffer (100 mM Tris-HCl, 100 mM EDTA, 1.5 M NaCl, 1% (w/v) CTAB (cetyltrimethylammonium bromide), 2% (w/v) SDS (sodium dodecyl sulfate), pH 8.0 and heated at 70 °C.

After 2 h, soil particulates are removed by centrifugation (30 min, 4000 x g). Crude eDNA is then precipitated from the resulting supernatant by adding 0.7 volumes of isopropanol and subsequently collected by centrifugation (30 min, 4000 x g). The resulting pellet is washed with 70% ethanol, air-dried and resuspended in a minimal volume of TE (10mM Tris, 1mM EDTA, pH 8.5). The remaining soil contaminants are removed by large-scale gel

purification on a 1% agarose gel (16 h, 20 V). Purified high-molecular-weight eDNA is recovered by electroelution (2 h, 100V), concentrated using centrifugal concentrators (MWCO 100 KDa) and end-repaired (End-It, Epicentre). Blunt-ended high molecular weight eDNA is ligated with a cosmid vector that has been digested with a blunt-end generating restriction enzyme and dephosphorylated. Ligation reactions are then packaged into lambda phage and the packaging reaction is used to transfect *E. coli*. Based on the number of colonies obtained from an initial test packaging reaction, this procedure should be scaled to yield at least 10,000,000 individual eDNA cosmid clones for an individual soil sample. We have found that soil eDNA cosmid libraries must contain at least 10,000,000 unique members (eDNA mega-libraries) before we are able to consistently recover overlapping clones for genetic loci of interest (Kim *et al.*, 2010).

2.2. Arraying and screening eDNA mega-libraries for overlapping clones

The procedure outlined below has been designed to facilitate the identification of clones in very large eDNA libraries. We have found that it is most effective to divide large libraries into multiple unique sub-libraries grouped into 8 × 8 or 8 × 12 grids. While sub-libraries of any size can be used we have found that pools of approximately 5,000 clones provide a good balance between the size of the arrayed library and the ability to efficiently recover individual clones of interest. A DNA miniprep and corresponding glycerol stock are created from each sub-library and these are used for PCR screening and clone recovery, respectively.

1. Based on the titer obtained in the small-scale titering experiment described above, set up an appropriate number of ligation and packaging reactions to obtain at least 250,000 eDNA clones. This will be sufficient to generate one 8 × 8 grid of sub-libraries containing 4,000–5,000 members each. We find it easiest to work with packaging reaction volumes that yield one grid at a time.
2. Confirm the titer of this large-scale reaction.
3. Carry out a large-scale phage infection reaction using the same conditions that were used for the small-scale titer reaction. Generally this will consist of mixing the large-scale packaging reaction with a 10-fold excess of an *E. coli* culture with an OD_{600} of 0.5–1.0.
4. After 1.5 h of shaking at 37 °C without any antibiotic selection, this mixture is used to inoculate 5 ml aliquots of LB (10g/l tryptone, 5g/l yeast extract, 10g/l NaCl, pH 7.0) (+ antibiotic) broth with enough of the transfection reaction to obtain approximately 5,000 unique transformants. Either 64 or 96 5-ml cultures should be set up depending on whether 8 × 8 or 8 × 12 grids are being used.
5. Incubate cultures overnight at 37 °C with shaking.
6. For each library pool remove 500 μl to create a glycerol stock and use the remainder for a DNA miniprep. These matching DNA miniprep and glycerol stock pairs will be used for PCR screening and clone recovery, respectively.
7. Array the DNA minipreps in either 8 × 8 or 8 × 12 grids.
8. Pool small aliquots (10μl) of DNA from each miniprep in a grid into a master aliquot for that grid. These “grid master pools” will be used for PCR screening. To facilitate higher-throughput screening, sub-libraries containing pools of aliquots from each row and/or column may also be generated and used as templates in the PCR screening reactions.

This process is repeated until the arrayed library contains at least 10,000,000 unique cosmid clones. Clones containing sequences of interest can be identified in the arrayed library using

the hierarchical pools miniprep DNA as templates in PCR reactions with primers designed to recognize sequences of interest. Once an amplicon-positive pool has been identified it is possible to use either colony hybridization or dilution PCR to recover the specific clone of interest from the corresponding glycerol stock (Pham *et al.*, 2007; Banik and Brady, 2008). The recovered cosmid clone is then end-sequenced using vector-specific primers and then primers designed to the end sequences are used to rescreen the arrayed library for overlapping clones. This process is continued until the full biosynthetic gene cluster is recovered from the library.

3. Use of TAR to reassemble overlapping clones into complete biosynthetic pathways

3.1 Capture vector construction

We developed the TAR capture vector pTARa to facilitate the reassembly of gene clusters in yeast and the subsequent introduction of these clusters into *Streptomyces* spp. by conjugation for heterologous expression studies (Feng *et al.*, 2010). In addition to the elements required for selection and propagation in yeast and *E. coli*, pTARa has been equipped with a *Dra*I fragment from pOJ436 (Bierman *et al.*, 1992) that contains an origin of transfer (*oriT*), an apramycin resistance marker and the phage ϕC31 integration system needed for integration into diverse *Streptomyces* spp. The protocol outlined below is written with pTARa in mind. Other capture vectors could be used with minor modifications to the general protocol.

For each TAR experiment, a capture vector must be outfitted with homology arms that correspond to sequences flanking the pathway to be reassembled. The cloning of the homology arms into the pTARa capture vector is outlined below:

1. Design two sets of primers, an upstream set (UPS1 and UPS2) that will amplify ~1kb of the proximal end of the left outermost clone and a downstream set (DWS1 and DWS2) that will amplify ~1 kb of the distal sequence of the right outermost clone (Table 1).
    a. The upstream forward primer (UPS1) should be designed to contain a *Bmt*I site (GCTAGC) for cloning into pTARa and therefore should read 5′-GCGC**GCTAGC** + 20 bp of the start of proximal targeting sequence-3′.
    b. The downstream reverse primer (DWS2) should be designed to contain an *Sph*I site for cloning into the pTARa vector and therefore should read 5′-GCGC**GCATGC** + 20 bp reverse complement of the end of the distal targeting sequence-3′.
    c. The downstream forward primer, DWS1, is the reverse complement of the upstream reverse primer, UPS2. Conceptually, DWS1 is the easier of the two to design and therefore its construction is described here. DSW1 consists of a 20-bp stretch of sequence that is ~1 kb downstream of UPS1 and a 20-bp stretch of sequence that is ~1 kb upstream of DWS2 (Figure 2). These two sequences are linked via an *Hpa*I site (**GTTAAC**) that will be used to linearize the capture vector prior to the TAR reaction.
2. Set up two first-round PCR reactions - one with primers UPS1 and UPS2 and the appropriate cosmid template and a second with primers DWS1 and DWS2 and the appropriate cosmid template.
3. Gel-purify the PCR products using a commercially available kit (e.g. Qiagen).

4. Set up a second round of PCR using primers UPS1 and DWS2 and 1 μl of each purified amplicon as a template. The first-round amplicons contain terminal complementary sequences that were introduced by the UPS2 and DWS1 primers and will therefore be linked together in the second round PCR reaction.
5. Gel-purify the resulting PCR product.
6. Digest the gel-purified amplicon with *BmtI/Sph*I and ligate this with similarly digested pTARa vector.
7. Introduce the ligation reaction by transformation into copy control *E. coli* (e.g. EPI300 Epicentre) and then check the newly created captured vector by *Hpa*I digestion to ensure you have the correct construct.
8. Use a copy control induction protocol (Epicentre) to obtain sufficient capture vector for the downstream TAR reaction. You will need 0.5 μg of digested capture vector for each TAR reaction.

3.2 Yeast spheroplast preparation

TAR requires the simultaneous co-transformation of yeast with the capture vector and the overlapping clones to be reassembled. Yeast transformation is accomplished by incubating DNA with spheroplasts, which are cells from which the cell wall has been almost completely removed. The following steps in the protocol outline the use of the yeast lytic enzyme, Zymolyase, to create the spheroplasts needed for carrying out the TAR reaction. This protocol has been modified from (Kouprina and Larionov, 2008).

1. Inoculate three uracil-deficient yeast CRY1-2 (MATa, ura3Δ, cyh2^R) colonies of different sizes in 50 ml YPD (Yeast extract Peptone Dextrose: 10 g/l yeast extract, 20 g/l peptone, 20 g/l glucose) for overnight growth at 30 °C.
2. Next day pick the culture with an OD_{660} of 3–5. This cell density is optimal for spheroplast preparation.
3. Harvest the cells by centrifugation at 1200 x g for 10 min at 4 °C.
4. Wash the pellet once by resuspending with 40 ml of sterile ddH_2O and spin as before.
5. Wash the cell pellet twice with 40 ml 1 M sorbitol as before.
6. Resuspend the washed pellet in 20 ml SCE (1 M sorbitol, 100 mM NaCitrate pH 5.8, 10 mM EDTA pH 8.0).
7. Add 40 μl of 2-mercaptoethanol along with 20 μl of a 10 mg/ml Zymolyase (Zymo Research) to the yeast suspension.
8. Incubate the mixture at 30 °C and monitor the spheroplast reaction every 20 min. Spheroplasting is monitored by diluting aliquots of the spheroplast mixture 1:10 in either 1 M sorbitol or 2% SDS and comparing the OD_{660} values of these solutions. Spheroplasts are protected by the sorbitol solution but will lyse in SDS.
9. When the ratio between intact (sorbitol aliquot) and lysed (SDS aliquot) spheroplasts reaches approximately 4, stop the reaction by centrifugation (500 x g, 5 min, 4 °C).
10. Remove the supernatant, gently resuspend the pellet in 20 ml of 1 M sorbitol (do not vortex) and then centrifuge as in step 9. Repeat washing procedure once. **[?]**
11. Resuspend the final spheroplast pellet in 2 ml STC (1 M sorbitol 10 mM Tris pH 7.5, 10 mM $CaCl_2$). Spheroplasts can be stored at room temperature for up to 1 h

before proceeding with the transformation protocol. Alternatively, spheroplasts can be aliquoted and stored at least up to one week at −80 °C.

3.3 Transformation of spheroplasts with pTARa capture vector along with overlapping clones capturing the entire biosynthetic pathway

Overlapping cosmid clones and the TAR vector must be linearized by restriction digestion prior to use in a TAR reaction. This section of the protocol outlines the restriction enzyme digestion of these vectors and the use of the linearized product to transform the yeast spheroplasts that were generated in the previous section of the protocol. Ultimately the transformation reactions are plated on uracil dropout plates to select for yeast that potentially contain successful TAR-reassembled gene clusters.

1. In a 20-μl reaction, digest 1 μg of each overlapping cosmid clone to be used in the assembly reaction with *Dra*I. Stop the reaction by heating at 70 °C for 20 minutes. The digested cosmids do not need to further purified.
2. Check an aliquot of the restriction digest on a 0.7% agarose gel to make sure *Dra*I alone does not cut in the eDNA insert. This gel should show a single fragment that co-migrates with the 23-kb band of a lambda *Hin*dIII marker. We have found *Dra*I to be a good enzyme to start with because it cuts in many commonly used cosmid and fosmid vectors (Supercos, pWEB, pWEB:TNC, etc.) and is rarely found in eDNA inserts containing secondary metabolite biosynthetic gene clusters because they often arise from GC-rich genomes such as those of actinobacteria like *Streptomyces*. If *Dra*I cuts in the eDNA insert other blunt-end cutting enzymes that cut in the vector should be explored for linearizing the clone.
3. Linearize 0.5 μg of pathway-specific capture vector with *Hpa*I.
4. Gel-purify the digested capture vector using a commercially available kit (Qiagen) according to the manufacturer's instructions.
5. Gently mix 200 ng of each linearized cosmid, an equimolar amount (~100 ng) of the linearized pathway specific-capture vector and 200 μl of spheroplasts.
6. Incubate at room temperature for 10 min and then add 800 μl of PEG solution (20% polyethylene glycol (PEG) 8000, 10 mM $CaCl_2$, 10 mM Tris pH 7.5).
7. Following a 10-minute incubation at room temperature, spin the transformation mix at 300 x g for 5 min at 4 °C.
8. Gently remove the supernatant and then resuspend the pellet in 800 μl SOS solution (1 M sorbitol, 6.5 mM $CaCl_2$, 0.25% yeast extract, 0.5% peptone).
9. Incubate the resuspended transformation mix at 30 °C for 40 min.
10. Add transformed spheroplasts to 7 ml of 50 °C SC top agar (1 M sorbitol, 1.92 g/l synthetic complete (SC) uracil dropout supplement, 6.7 g/l yeast nitrogen base, 2% glucose, 2.5% agar).
11. Overlay the spheroplast top agar mixture onto SC uracil dropout agar plates (same composition as above but 2% agar).
12. Incubate plates at 30 °C. Colonies typically appear within 72 h.

3.4 Analysis of TAR-recombined clones

In order to identify pTARa clones containing reassembled gene clusters individual yeast colonies are screened by PCR for the presence of fragments from each overlapping clone. Conveniently, the same primer pairs that were used to recover individual clones from the

library can also be used in this PCR screen. The initial PCR analysis should be coupled with comparative restriction mapping and if possible full BAC sequencing to confirm the integrity of the reassembled DNA construct. In our experience, approximately 20% of yeast transformants carry the correctly reassembled biosynthetic pathway.

1. Pick yeast transformants with toothpicks and patch them onto SC uracil dropout plates for overnight growth at 30 °C.
2. Resuspend a small portion of the patch in 10 μl of 20 mM NaOH and heat this mixture at 95 °C for 10 min.
3. Using 1.5 μl of the solution from step 2 as a template, PCR screen each transformant with primer pairs that recognize DNA fragments found in each of the overlapping clones. It is often possible to use the same primers here that were used in the clone recovery step.
4. Grow yeast recombinants that are PCR positive for all parts of a pathway overnight in 2 ml of SC uracil dropout medium (30 °C, 225 rpm).
5. Isolate the pTARa reassembled construct using ChargeSwitch™ Yeast Plasmid Isolation Kit (Invitrogen) per manufacturer's instructions.
6. Transform electrocompetent copy control *E. coli* (e.g. EPI300 from Epicentre) with 5 μl of the isolated pTARa construct and select on LB plates supplemented with 12.5 μg/ml chloramphenicol.
7. On the next day inoculate a single transformant into 10 ml of LB with 12.5 μg/ml chloramphenicol supplemented with 60 μl of copy control BAC autoinduction solution (Epicentre).
8. Incubate at 37 °C for 20 h.
9. Isolate the reassembled pathway-containing BAC using alkaline lysis and then purify the DNA by phenol/chloroform extraction and isopropanol precipitation.
10. Digest the BAC clone using an enzyme that yields well-resolved DNA fragments by agarose gel electrophoresis. If the sequence data for the pathway of interest is available compare the experimental restriction band pattern of the assembled BAC clone with the theoretically predicted band sizes of the clone. If the full sequence is not available, digest all overlapping clones with the same restriction enzyme and compare the restriction patterns produced by the reassembled and the individual clones. It may be necessary to use a 0.5% agarose gel run at 30 V overnight to resolve large restriction fragments.

4. Downstream analysis

The ultimate aim of this protocol is to transfer faithfully reassembled gene clusters into model cultured bacterial hosts for heterologous expression studies in order to gain functional access to the metabolites encoded by these gene clusters. pTARa was specifically designed for use in *Streptomyces* and can be transferred into these bacteria by intergeneric conjugation using *E. coli* ET12567/pUZ8002 (Kan^R/Cm^R) as a donor strain following standard protocols (Kieser *et al.*, 2000). Organic extracts from cultures of *Streptomyces* spp. transformed with either the individual cosmid clones or the reassembled pathway can be compared by LCMS to look for the heterologous production of pathway-encoded metabolites. By exchanging the ϕC31 integrase system for any of a number of broad-host-range origins of replication it could be used in a phylogenetically diverse collection of model cultured bacterial hosts.

5. Example

Figure 3 outlines an example where this protocol was used to identify novel metabolites encoded by an eDNA gene cluster captured on two overlapping clones (Feng *et al.*, 2010). Degenerate primers based on conserved regions of minimal polyketide synthase (PKS) genes were used to amplify full-length ketosynthase$_\alpha$ genes (those encoding the alpha subunit of the heterodimeric Type II ketosynthase) from DNA minipreps in an arrayed eDNA library. Many of these genes exhibited low identity to known ketosynthase$_\alpha$ sequences, suggesting that they were likely associated with gene clusters distinct from previously sequenced clusters and might therefore encode the biosynthesis of novel secondary metabolites. In the example outlined in Figure 3, extracts from cultures of *Streptomyces albus* transformed with the eDNA-derived Type II PKS-containing clone AB649 were found to contain the tetracyclic polyketide intermediates rabelomycin and dehydrorabelomycin. Complete sequencing of AB649 indicated that the biosynthetic system responsible for the production of these metabolites likely extended beyond the sequence captured on this clone. DNA minipreps from the arrayed eDNA library were rescreened with primers designed to recognize the terminal eDNA sequence captured in AB649 and the overlapping clone AB1850 was recovered in this screen. Using TAR, AB649 and AB1850 were reassembled into a large-insert BAC (AB649/1850) that was predicted to contain a complete biosynthetic gene cluster. By PCR analysis, restriction digestion and full sequencing this reassembled construct was found to be a faithful reassembly of the overlapping eDNA clones. Isolation and structure elucidation of metabolites obtained through heterologous expression of this intact gene cluster identified three novel fluostatins. The development of TAR as a tool for metagenomics makes it possible to routinely clone and functionally study even very large gene clusters directly from soil microbiomes.

Acknowledgments

This work was supported by NIH GM077516. SFB is a Howard Hughes Medical Institute early career scientist.

References

Banik JJ, Brady SF. Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. Proc Natl Acad Sci U S A. 2008; 105:17273–17277. [PubMed: 18987322]

Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature. 2002; 417:141–147. [PubMed: 12000953]

Bierman M, Logan R, O'Brien K, Seno ET, Rao RN, Schoner BE. Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. Gene. 1992; 116:43–49. [PubMed: 1628843]

Binz TM, Wenzel SC, Schnell HJ, Bechthold A, Muller R. Heterologous expression and genetic engineering of the phenalinolactone biosynthetic gene cluster by using red/ET recombineering. Chembiochem. 2008; 9:447–454. [PubMed: 18157854]

Brady SF. Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. Nature Protocols. 2007; 2:1297–1305.

Brady SF, Simmons L, Kim JH, Schmidt EW. Metagenomic approaches to natural products from free-living and symbiotic organisms. Nat Prod Rep. 2009; 26:1488–1503. [PubMed: 19844642]

Daniel R. The metagenomics of soil. Nat Rev Microbiol. 2005; 3:470–478. [PubMed: 15931165]

Feng Z, Kim JH, Brady SF. Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived type II PKS gene cluster. J Am Chem Soc. 2010; 132:11902–11903. [PubMed: 20690632]

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 3rd, Smith HO. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science. 2008a; 319:1215–1220. [PubMed: 18218864]

Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA 3rd. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. Proc Natl Acad Sci U S A. 2008b; 105:20404–20409. [PubMed: 19073939]

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol. 1998; 5:R245–249. [PubMed: 9818143]

Kieser, T.; Bibb, MJ.; Buttner, MJ.; Chater, KF.; Hopwood, DA. Practical Streptomyces genetics. The John Innes Foundation; Norwich, United Kingdom: 2000.

Kim JH, Feng Z, Bauer JD, Kallifidas D, Calle PY, Brady SF. Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. Biopolymers. 2010; 93:833–844. [PubMed: 20577994]

Kouprina N, Larionov V. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae. Nat Protoc. 2008; 3:371–377. [PubMed: 18323808]

Kouprina N, Noskov VN, Larionov V. Selective isolation of large chromosomal regions by transformation-associated recombination cloning for structural and functional analysis of mammalian genomes. Methods in molecular biology. 2006; 349:85–101. [PubMed: 17071976]

Larionov V, Kouprina N, Eldarov M, Perkins E, Porter G, Resnick MA. Transformation-associated recombination between diverged and homologous DNA repeats is induced by strand breaks. Yeast. 1994; 10:93–104. [PubMed: 8203155]

Larionov V, Kouprina N, Graves J, Chen XN, Korenberg JR, Resnick MA. Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. Proc Natl Acad Sci U S A. 1996; 93:491–496. [PubMed: 8552668]

Mathee K, Narasimhan G, Valdes C, Qiu X, Matewish JM, Koehrsen M, Rokas A, Yandava CN, Engels R, Zeng E, Olavarietta R, Doud M, Smith RS, Montgomery P, White JR, Godfrey PA, Kodira C, Birren B, Galagan JE, Lory S. Dynamics of Pseudomonas aeruginosa genome evolution. Proc Natl Acad Sci U S A. 2008; 105:3100–3105. [PubMed: 18287045]

Nagano Y, Takao S, Kudo T, Iizasa E, Anai T. Yeast-based recombineering of DNA fragments into plant transformation vectors by one-step transformation. Plant cell reports. 2007; 26:2111–2117. [PubMed: 17680244]

Pham VD, Palden T, DeLong EF. Large-scale screens of metagenomic libraries. J Vis Exp. 2007; 201

Rappe MS, Giovannoni SJ. The uncultured microbial majority. Annu Rev Microbiol. 2003; 57:369–394. [PubMed: 14527284]

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J, Goodman RM. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol. 2000; 66:2541–2547. [PubMed: 10831436]

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]

Simon C, Daniel R. Achievements and new knowledge unraveled by metagenomic approaches. Appl Microbiol Biotechnol. 2009; 85:265–276. [PubMed: 19760178]

Simon C, Daniel R. Construction of small-insert and large-insert metagenomic libraries. Methods Mol Biol. 2010; 668:39–50. [PubMed: 20830554]

Torsvik V, Ovreas L, Thingstad TF. Prokaryotic diversity--magnitude, dynamics, and controlling factors. Science. 2002; 296:1064–1066. [PubMed: 12004116]

Torsvik V, Salte K, Sorheim R, Goksoyr J. Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. Appl Environ Microbiol. 1990; 56:776–781. [PubMed: 2180371]

Zhou J, Bruns MA, Tiedje JM. DNA recovery from soils of diverse composition. Appl Environ Microbiol. 1996; 62:316–322. [PubMed: 8593035]

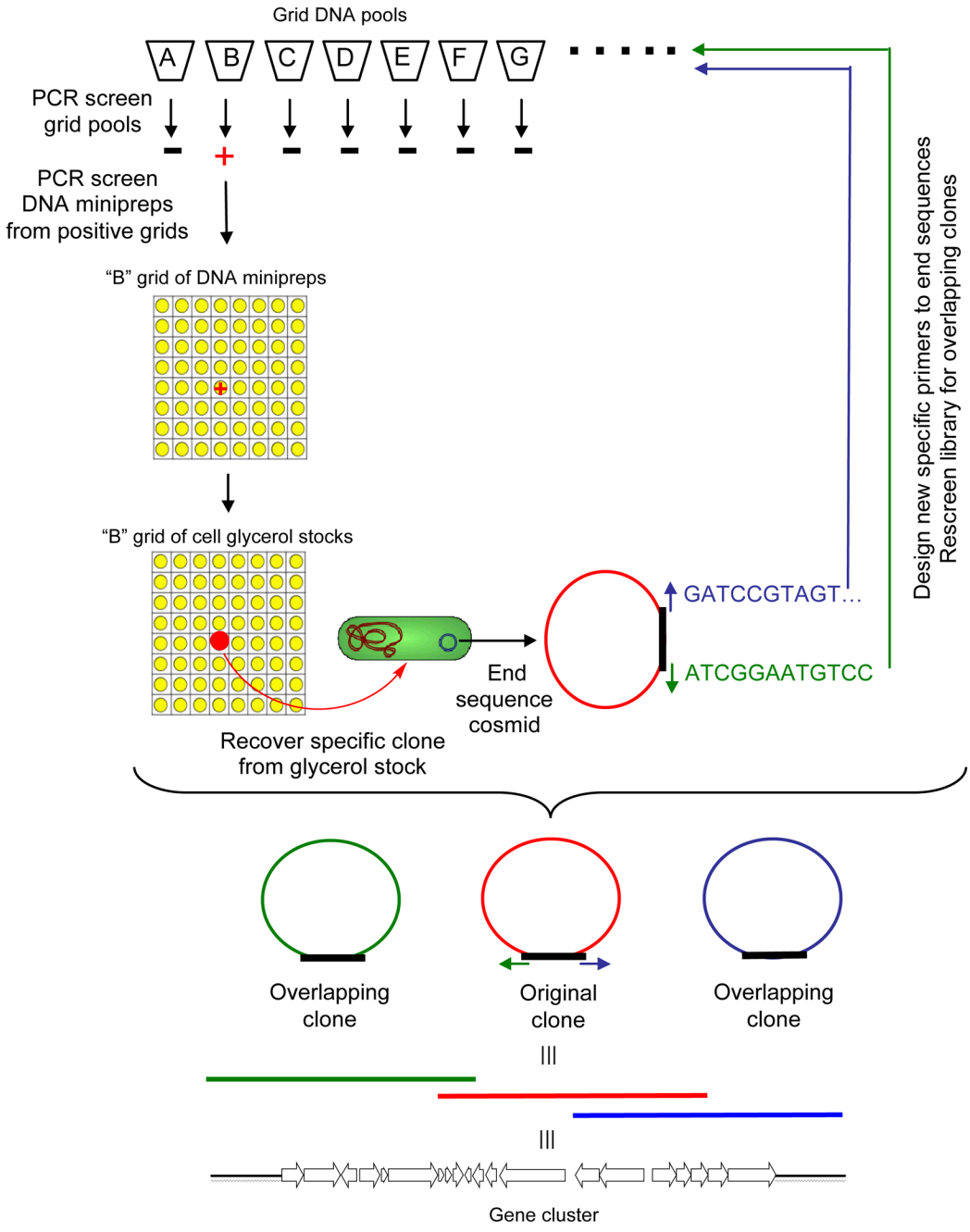

Figure 1.

PCR screening of an arrayed environmental DNA library. Once an initial clone of interest is recovered, new primers are designed based on the end sequencing of this clone. These new primers are then used to rescreen the library for overlapping clones. The screening cycle is repeated until the full biosynthetic pathway is recovered.

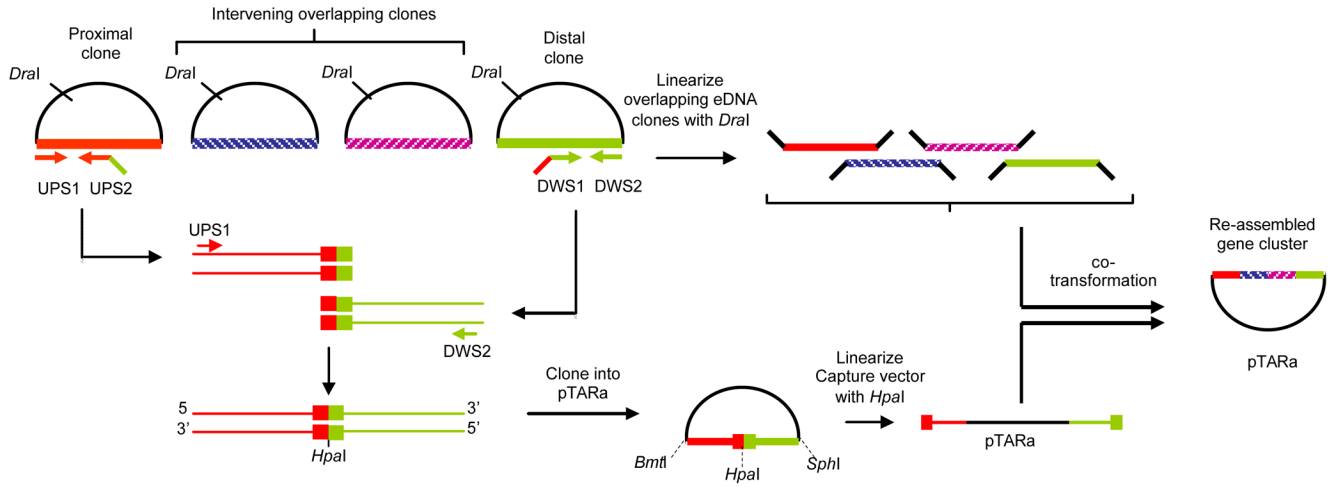

Figure 2.

TAR reassembly of gene clusters capture on overlapping cosmid clones. The proximal and the distal ends of the outermost clones are amplified using upstream and downstream sets of primers (UPS1/UPS2, DWS1/DWS2), respectively. Primers UPS2 and DWS1 are reverse complement sequences and therefore these two amplicons can be linked with a second round of PCR. The resulting amplicons, consisting of clone specific homology arms is ligated into pTARa capture vector. Linearized capture vector (*Hpa*I) and linearized overlapping clones (*Dra*I) are co-transformed in yeast for the recombination reaction.

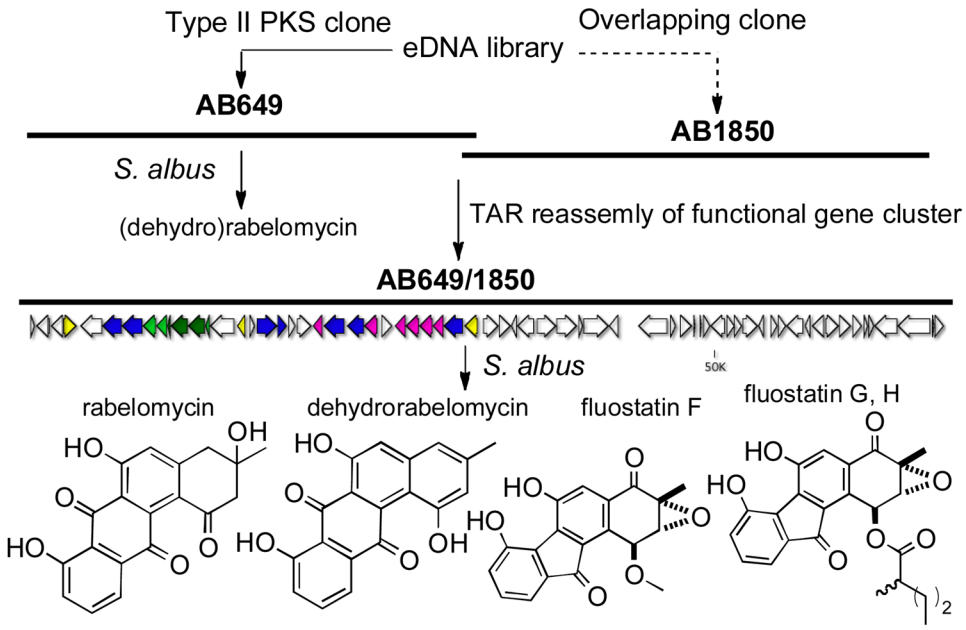

Figure 3.

Fluostatins F, G and H were isolated from cultures of *S. albus* transformed with a TAR reassembled gene cluster captured on eDNA clones AB649 and AB1850. *S. albus* transformed with clone AB649 alone produces only the intermediates rabelomycin and dehydrorabelomycin. *S. albus* transformed with clone AB1950 produced no detectable clone specific secondary metabolites.

Table 1

Primers used for capture vector construction.

Name	Primer description
UPS1	5′-GCGC**GCTAGC** + 20 bp of the start of proximal targeting sequence-3′[a]
UPS2	Reverse complement of DWS1
DWS1	5′-20 bp that is ~1 kb downstream of UPS1 + **GTTAAC** + 20 bp that is ~1 kb upstream of DWS2-3′[b]
DWS2	5′-GCGC**GCATGC** + 20 bp reverse complement of the end of the distal targeting sequence-3′[c]

The added *Bmt*I[a], *Hpa*I[b] and *Sph*I[c] sites are shown in bold.