FEMS Microbiology Letters, 365, 2018, fny064

doi: 10.1093/femsle/fny064
Advance Access Publication Date: 14 March 2018
Minireview

MINIREVIEW – Biotechnology & Synthetic Biology

Large inserts for big data: artificial chromosomes in the genomic era

Arianna Tocchetti[1], Stefano Donadio[1,2] and Margherita Sosio[1,2,*]

[1]Naicons Srl, 20139 Milano, Italy and [2]Ktedogen Srl, Via Ortles 22/4, 20139 Milano, Italy

*Corresponding author: Naicons Srl, Via Ortles 22/4, 20139 Milano, Italy. E-mail: msosio@naicons.com

One sentence summary: Vectors based on bacterial or P1-derived artificial chromosomes, on transformation-associated and on integrase-mediated recombination, developed for capturing and heterologous expression of large biosynthetic gene clusters for new bioactive compound discovery and production.

Editor: Marnix Medema

ABSTRACT

The exponential increase in available microbial genome sequences coupled with predictive bioinformatic tools is underscoring the genetic capacity of bacteria to produce an unexpected large number of specialized bioactive compounds. Since most of the biosynthetic gene clusters (BGCs) present in microbial genomes are cryptic, i.e. not expressed under laboratory conditions, a variety of cloning systems and vectors have been devised to harbor DNA fragments large enough to carry entire BGCs and to allow their transfer in suitable heterologous hosts. This minireview provides an overview of the vectors and approaches that have been developed for cloning large BGCs, and successful examples of heterologous expression.

Keywords: large biosynthetic gene cluster; *E. coli-Streptomyces* shuttle artificial chromosomes; pESAC; TAR system; IR system

INTRODUCTION

Bioactive compounds of microbial origin with their structural novelty, diversity and complexity are still an invaluable source for drug discovery and development (Monciardini *et al.* 2014; Genilloud 2017). The exponential increase in available microbial genome sequences coupled with predictive bioinformatic tools are underscoring the genetic capacity of bacteria to produce an unexpected large number of specialized compounds. However, only a small fraction of these compounds can be detected in the laboratory, since most of the corresponding biosynthetic genes are silent or poorly expressed under conventional laboratory culture conditions (Rutledge and Challis 2015; Blin *et al.* 2017; Tracanna *et al.* 2017). While the organization of genes in biosynthetic gene clusters (BGCs) facilitates their identification and engineering, these approaches are often hampered by lack of or limited genetic tractability of the producing microrganisms (Galm and Shen 2006; Rutlege and Challis 2015). Thus, standard approaches such as overexpression of a cluster-associated activator, deletion of a cluster-associated repressor, replacement of natural promoters with constitutive or inducible promoters, or overexpression of pathway precursors are not immediately applicable to most microbial strains. Devising methods to genetically manipulate a microbial strain with interesting BGC(s) can be a lengthy and sometime unsuccessful endeavor. This is particularly true for actinomycetes, the main producers of secondary metabolites and the focus of this minireview.

Important research tools for genome projects were developed in the 1990s, when DNA sequencing heavily depended on producing high-quality libraries, particularly large-insert libraries, well before the introduction of next-generation sequencing (NGS) and the deluge of microbial genome sequences. Relevant tools were yeast artificial chromosomes (YACs) for cloning and assembling large DNA fragments (Monaco and Larin 1994) and artificial chrosomomes for replication in bacteria (Shizuya *et al.* 1992; Ioannou *et al.* 1994). The latters were distinguished on the basis of the replicon they employed and named

Received: 9 January 2018; **Accepted:** 13 March 2018

BAC and PAC if they relied on plasmid F1- or bacteriophage P1-derived replicon, respectively. YAC, BAC and PAC vectors could easily accommodate DNA inserts exceeding 100 kbp.

The logical step was then to combine BAC or PAC vectors with genetic elements that allowed their introduction in suitable actinomycete hosts. One of the early examples is represented by the PAC-derived *Escherichia coli-Streptomyces* shuttle vectors, designated ESACs, constructed for antibiotic-producing actinomycetes (Sosio *et al.* 2000; Miao *et al.* 2005; Galm and Shen 2006). ESAC vectors could assist in various genetic manipulations of the biosynthetic pathways through cloning of entire clusters and their heterologous expression in a host amenable to *in vivo* genetic manipulation. However, at the time ESAC and related vectors were developed, rapid whole genome sequencing was not available, and methods for creating large-insert DNA libraries were not well developed; thus, the use of ESAC vectors could require assembling large DNA inserts from cosmids through multiple rounds of homologous recombination (Sosio, Bossi and Donadio 2001).

While YAC, BAC and PAC tools are no longer necessary with NGS technology, their use is witnessing a renaissance thanks to the increasing number of deposited genome sequences, improved bioinformatics tools for BGC identification and improved methods for library construction. Although the successful heterologous expression of an entire BGC remains unpredictable, several cloning and heterologous host systems have been developed, providing ample choices (Alduina and Gallo 2012; Ongley *et al.* 2013; Nah *et al.* 2017).

In this minireview, we will provide an overview of the vectors that have been developed for cloning large BGC and the strategy that has been used for their successful heterologous expression. Because of space constraints, this overview will be limited to actinomycete-derived metabolites. Nonetheless, many of the principles developed for actinomycetes have been successfully employed for other important hosts such as *E. coli*, *Pseudomonas putida*, *Bacillus subtilis* (Ongley *et al.* 2013; Rutledge and Challis 2015; Nah *et al.* 2017).

POSSIBLE APPROACHES FOR HETEROLOGOUS EXPRESSION

Heterologous expression of an entire BGC represents one of the most promising approaches to overcome the hurdle of developing a gene transfer system for the native producer. It not only allows the production of a desired metabolite in a host amenable to genetic manipulation, but it can also lead to the development of general rules for efficient production of different classes of metabolites in defined host(s), without the need to devise *ad hoc* conditions for cultivation of or establish metabolic models in different hosts (Ongley *et al.* 2013).

Vectors suited for heterologous expression of entire BGCs must be effectively mobilized into the chosen host(s), where they can be stably maintained, possibly without antibiotic addition to select for construct presence. Normally, this is achieved by site-specific integration of the incoming DNA by a vector-encoded integrase. In addition, the ability to accommodate large DNA inserts is an important requisite, since many BGCs are too large to fit within the cloning capacity of cosmid and fosmid vectors. Various approaches for heterologous expression of secondary metabolite biosynthetic pathway have been pursued in the last few years, and they made use of different *E. coli-Streptomyces* shuttle vectors based on bacterial or P1-derived artificial chromosomes (Jones *et al.* 2013; Alt and Wilkinson 2015; Nah *et al.* 2015; Kepplinger *et al.* 2017; Pyeon *et al.* 2017), of transformation-associated recombination (TAR; Yamanaka *et al.* 2014) and of integrase-mediated recombination (IR; Du *et al.* 2015). These approaches are schematized in Fig. 1. Table 1 lists the key features of the BAC, PAC, TAR and IR vectors used over the last decade for cloning and heterologous expression. Figure 2 shows examples of secondary metabolites heterologously expressed with the herein described approaches.

CLONING SYSTEMS FOR LARGE BGCs

The TAR system

One of the early reports made use of RecET-mediated linear-plus-linear homologous recombination and was used successfully to clone BGCs of 10–52 kbp from the genome of *Photorhabdus luminescens* and express two of them in *E. coli* (Fu *et al.* 2012; Ongley *et al.* 2013). Based on this system, TAR was used for direct cloning of 20–73 kbp BGCs (Yamanaka *et al.* 2014). The TAR system exploits the high level of homologous recombination in the yeast *Saccharomyces cerevisiae*. Recent TAR cloning vectors, such as pCAP01, consist of three elements: for yeast, an autonomous replicating sequence (ARS), a plasmid maintenance element (CEN) and an auxotrophic marker; for *E. coli*, a pUC *ori* that could stably carry inserts larger than 50 kbp and an *oriT* for conjugation; and for actinobacteria, an *attP-int* system for chromosomal integration (Fig. 3). The capture vectors are also provided with an antibiotic resistance gene for selection of both *E. coli* and actinomycete recombinants (Yamanaka *et al.* 2014; Tang *et al.* 2015; Bilyk *et al.* 2016). In this way, the target DNA can be assembled in yeast, isolated and introduced into *E. coli*, and then directly conjugated into the desired actinomycete, where it can be maintained after chromosomal integration (Table 1). The TAR strategy relies on the 'capturing vector', which contains the two distal ends of the target DNA segment. The linearized capturing vector and the restriction enzyme digested genomic DNA are assembled into yeast, where the target DNA fragment is captured by homologous recombination (Fig. 1B).

Examples of BGCs isolated by the TAR technique using pCAP01 vector and expressed in a heterologous host include (Table 2) the 73-kbp taromycin A BGC from a marine *Saccharomonospora* sp. (Fig. 2), which was successfully expressed in *Streptomyces coelicolor* M1146 after deleting the gene for a LuxR-type regulator (Yamanaka *et al.* 2014); the 21-kbp enterocin BGC from *Salinispora pacifica*, which was expressed in two different *Streptomyces* strains (Bonet *et al.* 2015); and the 44-kbp nataxazole BGC from *Streptomyces* sp. TÜ6176 which led to production of the nataxazole precursor AJ9561 in *S. lividans* JT46 (Cano-Prieto *et al.* 2015). In addition, Jordan and Moore (2016) identified the ammosamide BGC through direct cloning in pCAP01 and heterologous expression in *S. coelicolor* M512, demonstrating the complex set of biosynthetic genes necessary for biosynthesis of this pyrroloquinoline alkaloid through gene deletions in the heterologous host (Table 2). A modified version of pCAP01 has been utilized to clone PKS-NRPS hybrid BGCs from two *Salinispora* strains and to identify thiotetronic acid antibiotics as their products after expression in *S. coelicolor* M1152 (Tang *et al.* 2015) (Fig. 2). The vector pCLY10, another TAR shuttle vector, was utilized to clone the 36-kbp grecocycline BGC from *Streptomyces* sp. Acta 1362 and to express it in *Staphylococcus albus* J1074, where several grecocycline congeners were detected (Bilyk *et al.* 2016).

An *in vitro* variation of the TAR system is represented by CATCH (Cas9-assisted targeting of chromosome). The method is

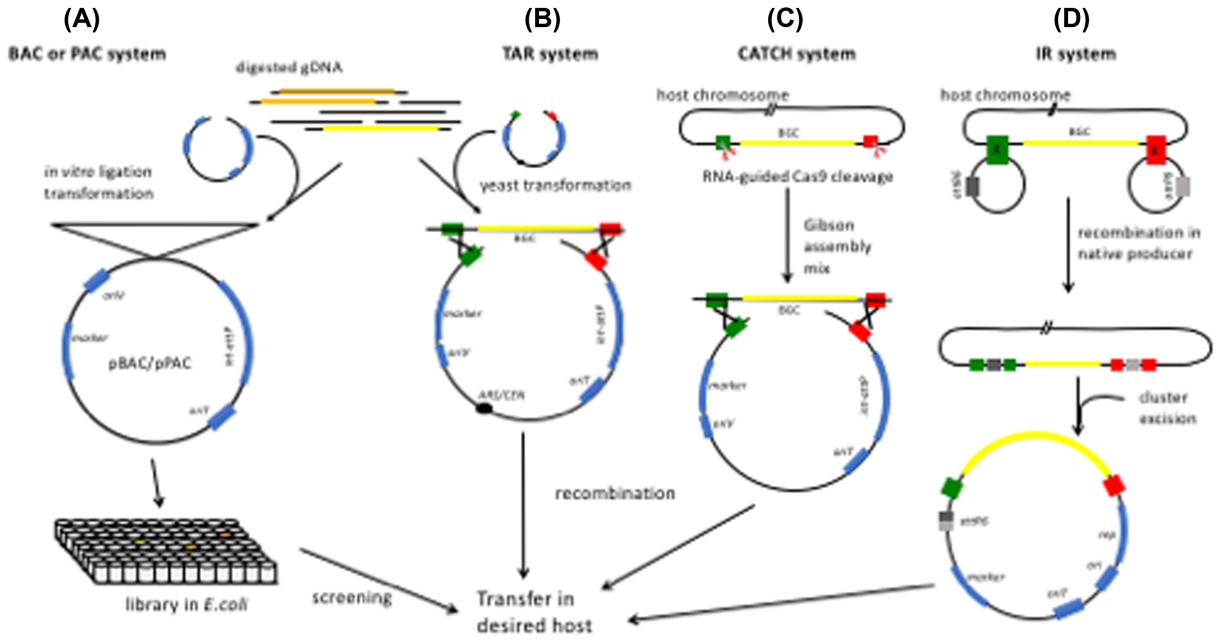

Figure 1. Schematic representations of different cloning systems for large BGCs. (**A**) Diagram of *in vitro* cloning using BAC/PAC vectors; (**B**) *in vivo* cloning using the transformation-associated recombination (TAR) system in yeast; (**C**) *in vitro* Cas9-Assisted Targeting of Chromosome (CATCH) system; (**D**) integrase-mediated recombination (IR) system in the native producer. BGCs are indicated by thick, colored lines. See text for details.

Table 1. List of cloning vectors for large inserts.

Cloning system	Vector	Size (kbp)	Replicon (*E. coli*)	Features	Derived from	References
BAC/PAC	pStreptoBAC V	16.2	F factor	*aac(3)IV*, *oriT-attP-int*$^{\phi C31}$	pBACe3.68 (Frengen *et al.* 1999)	Miao *et al.* 2005
	pSBAC	12.0	*ori2* (single copy) *oriV* (high copy)	*aac(3)IV*, *oriT-attP-int*$^{\phi BT1}$ *parA*, *parB* and *par C*	pCC1BAC (EPICENTRE)	Liu *et al.* 2009
	pESAC13	23.3	P1 lytic replicon	*aph(3)II*, *tsr*, *oriT-attP-int*$^{\phi C31}$	pPAC-S1 (Sosio *et al.* 2000)	www.biost.com; Jones *et al.* 2013
	pESAC13A	22.7	P1 lytic replicon	*aac(3)IV*, *tsr*, *oriT-attP-int*$^{\phi C31}$	pESAC13	This work
TAR	pCAP01	9.0	pUC *ori*	*aph(3)II*, *oriT-attP-int*$^{\phi C31}$ *TRP1* *ARSH4/CEN6*	SuperCos1 (Stevens, Hari and Boddy 2013)	Yamanaka *et al.* 2014
IR	pSV::*attB6*-HR' + pKC1139::*attP6*-HR"		–	*neo*, *oriT*, *attB*	pUC119::*neo* (Li *et al.* 2009)	Du *et al.* 2015
			pUC *ori*	*aac(3)IV*, *oriT*, *ori*^{ts} + *rep attP6*	pKC1139 (Bierman *et al.* 1992)	

aac(3)IV, apramycin resistance marker; *aph(3)II*, kanamycin/neomycin resistance marker; *neo*, kanamycin resistance marker; *tsr*, thiostrepton resistance marker; *TRP1*, auxotrophic marker; *int*$^{\phi C31}$, ϕC31 integrase; *int*$^{\phi BT1}$, ϕBT1 integrase; *oriT*, origin of transfer; ARSH4/CEN6, yeast origin of replication and centromere; *ori*^{ts} + *rep*, temperature sensitive origin and replication protein; *attB6*, modified attachment site; *attP6*, modified attachment site; *parA*, *parB* and *parC*, *E. coli* partitioning system genes; HR', HR", homologous regions.

based on the *in vitro* application of RNA-guided Cas9 nuclease (Jiang and Zhu 2016), which can cleave *in vitro* intact bacterial chromosomes embedded in agarose plugs at target sequences. The resulting fragments can be subsequently ligated through Gibson assembly (Gibson *et al.* 2009) (Fig. 1C). This system was used to clone the 36-kbp jadomycin BGC from *S. venezuelae* and the 32-kbp chlortetracycline BGC from *S. aureofaciens* (Jiang *et al.* 2015).

The IR system

Recently, a strategy was developed based on integrase-mediated site-specific recombination and used for simultaneously engineering a *Streptomyces* genome and cloning a BGC. The *attP-int* loci from phages ΦC31 and ΦBT1 have been used to construct versatile vectors which can integrate into different *attB* sites in *Streptomyces* (Bierman *et al.* 1992; Gregory, Till and Smith 2003; Gonzalez-Quiñonez *et al.* 2016). To increase the diversity of the

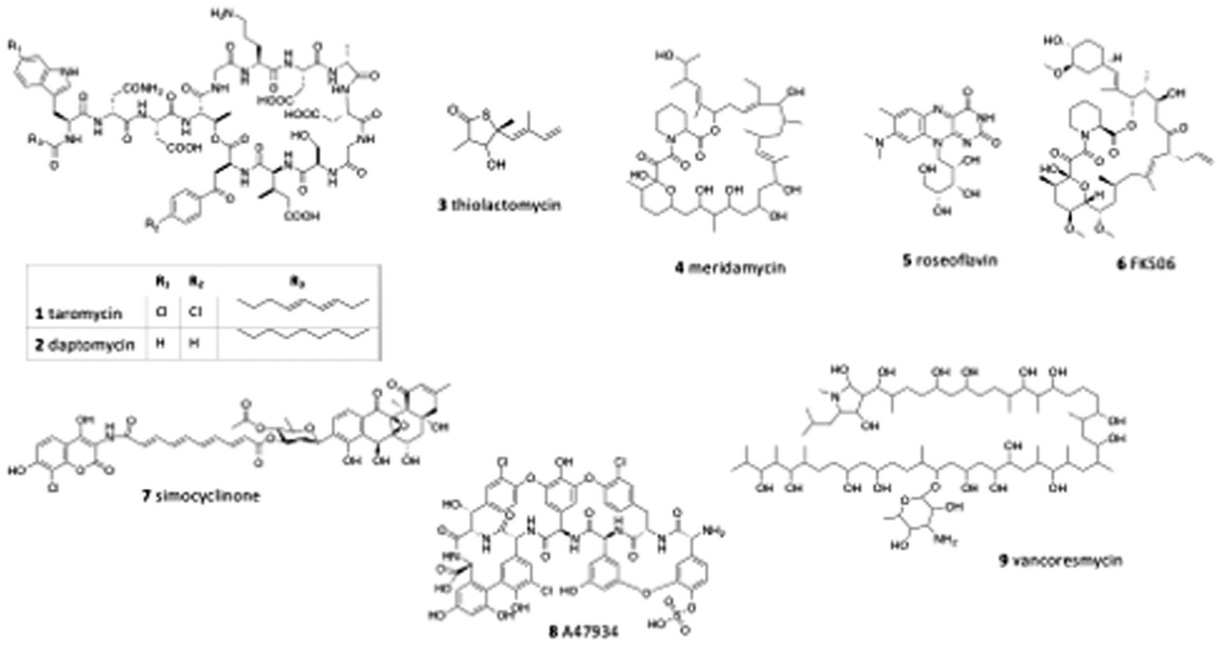

Figure 2. Examples of metabolites produced by heterologous expression using the TAR system (1, 3) or BAC/PAC vectors (2, 4–8).

ΦBT1 *attP-attB* pair, the central dinucleotide sequence of *attB* and *attP* was mutated (Du *et al.* 2015). Using a non-replicating and a temperature-sensitive replicon, homologous recombination was used for targeted integration of the mutated *attB* and *attP* at the distal ends of the BGC of interest. Then, ΦBT1 Int excised the target region from the chromosome, and the plasmid containing the BGC was recovered at the permissive temperature (Fig. 1D). This system was validated by cloning the 25-kbp actinorhodin (*act*) BGC, the 45-kbp napsamycin BGC and a 157-kbp segment containing the daptomycin BGC (Fig. 2). The fidelity of the procedure was confirmed by complementation of a Δ*act S. coelicolor* mutant (Du *et al.* 2015).

Escherichia coli-Streptomyces shuttle BAC/PAC vectors

Pioneering work with BAC and PAC vectors demonstrated the capacity to clone DNA inserts approaching 200 kb in length, which could be stably maintained in a heterologous *Streptomyces* host (Sosio *et al.* 2000; Martinez *et al.* 2004; Miao *et al.* 2005). The early *E. coli-Streptomyces* shuttle vectors pBELOBAC and pPAC-S1 have since been replaced by several improved variants, such as pStreptoBAC V, pSBAC and pESAC13, which contains the *oriT* site for conjugative transfer into actinomycetes and have been successfully used for heterologous expression (Table 1, Table 3 and Fig. 3).

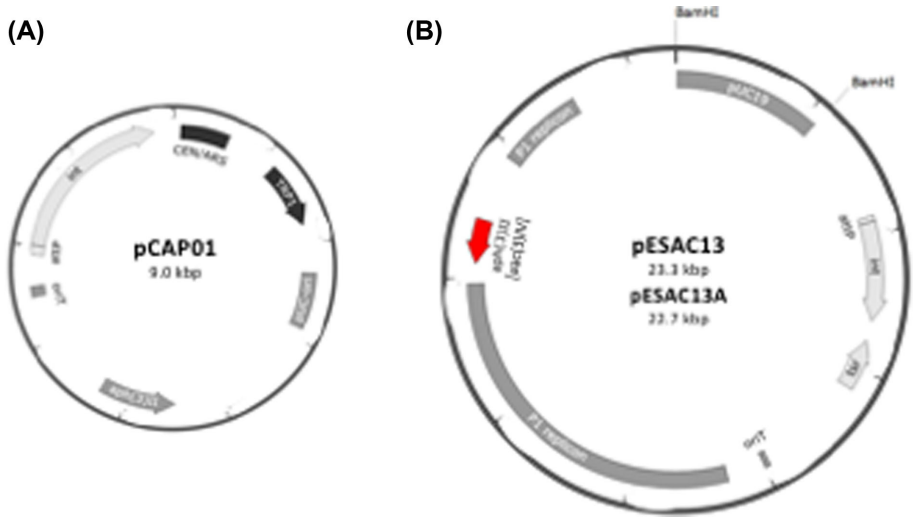

Figure 3. Circular maps of pCAP01, pESAC13 and pESAC13A. (**A**) pCAP01 consists of three elements: for direct cloning and manipulation in yeast (CEN/ARS, TRP1); for maintenance and manipulation in *E. coli* (pUC *ori*) and for chromosomal integration and expression in actinobacteria (*attP-int* system) (Yamanaka *et al.* 2014); (**B**) pESAC13 was constructed by inserting a 760-bp DNA fragment containing *oriT* from the RK2 replicon (Simon, Priefer and Pühler 1983) into the unique *Bst*XI site of pPAC-S1 (Sosio *et al.* 2000). pESAC13A was constructed by replacing through Red/ET recombineering the *aph(3)II* gene (conferring kanamycin resistance) in pESAC13 with the *aac(3)IV* gene (conferring apramycin resistance) from pSET152 (Bierman *et al.* 1992).

Table 2. Examples of heterologous expression of BGCs cloned by the TAR approach.

Vectors	Natural product (producer)	Biosynthetic class	BGC size (kbp)	Heterologous host	References
pCAP01	Marinopyrrole (*Streptomyces* sp. CNQ418)	NRP/PK	30	*S. coelicolor* M512	Yamanaka *et al.* 2014
	Taromycin A (*Saccharomonospora* sp. CNQ-490)	NRP	73	*S. coelicolor* M1146[a]	Yamanaka *et al.* 2014
	Enterocin (*Salinispora pacifica* CNT-150)	PK	21	*S. lividans* TK23 *S. coelicolor* M1146	Bonet *et al.* 2015
	Nataxazole (*Streptomyces* sp. TÜ6176)	PK	44	*S. lividans* JT46	Cano-Prieto *et al.* 2015
	Ammosamides A-C (*Streptomyces* sp. CNR-698)	RiPP	35	*S. coelicolor* M512	Jordan and Moore 2016
pCAP03[b]	Thiolactomycin (*S. pacifica CNS-863*)	PK-NRP	22	*S. coelicolor* M1152	Tang *et al.* 2015
	Tiotetroamide (*S. afghaniensis* NRRL 5621)	PK-NRP	33	*S. coelicolor* M1152	Tang *et al.* 2015
pCLY10[c]	Grecocycline (*Streptomyces* sp. Acta 1362)	PK	36	*St. albus* J1074	Bilyk *et al.* 2016

PK = polyketide; NRP = non-ribosomal peptide; RiPP = ribosomally synthesized, posttranslationally modified peptide;
[a]after cluster refactoring;
[b]derivative of pCAP01 characterized by the counterselectable marker *URA3* under the strong pADH1 promoter; see Tang *et al.* (2015) for further details;
[c]shuttle vector for *E.coli*-yeast-actinomycetes (Bilyk *et al.* 2016).

The vector pStreptoBAC V, which contains the ΦC31 *attP-int* system (Table 1), was used for cloning the 128-kb segment containing the BGC for the NRPS-made lipopeptide daptomycin (Fig. 2) and for its analog A21978C from *S. roseosporus* NRRL 11379, and their expression in different *S. lividans* strains (Miao *et al.* 2005; Penn *et al.* 2006). Although the antibiotic yields were low, production was improved by deleting the *act* BGC and by medium optimization (Penn *et al.* 2006). Later, pStreptoBAC V was used for cloning and heterologous expression of the 65-kb iso-migrastatin BGC in different *Streptomyces* hosts (Feng *et al.* 2009).

The pSBAC vector, which contains the ΦBT1 *attP-int* system (Table 1), was used for cloning the 90-kb PKS-NRPS BGC for the macrolide meridamycin (Fig. 2) from *Streptomyces* sp. NRRL 30748 and its heterologous expression in *S. lividans* K4-114, a Δ*act* strain. Production of meridamycin and its 3-nor analog was observed only upon inserting the *ermE∗* promoter in front of the NRPS-PKS genes, and it was enhanced by feeding the biosynthetic precursor diethylmalonate (Liu *et al.* 2009). The pSBAC vector has been subsequently used for cloning two PKS I BGCs, for tautomycetin (80 kb from *Streptomyces* sp. CK4412) and pikromycin (60 kb from *S. venezuelae* ATCC 15439), applying the plasmid rescue technique (Nah *et al.* 2015; Pyeon *et al.* 2017). For tautomycetin, while introduction of the pSBAC in naïve strains led to similar production levels as in the original host, tandem transfer of the constructs in *Streptomyces* sp. CK4412 or in *S. coelicolor* M145 dramatically increased tautomycetin production (Nah *et al.* 2015). Similarly, when the pikromycin BGC was tandemly introduced into *S. lividans* TK21, production of 10-deoxymethynolide and pikromycin was observed, at levels 1.6- to 1.8-fold the parental strain (Pyeon *et al.* 2017).

Over the last few years, there have been several reports of heterologous expression of a variety of natural products, mostly generated by large modular PKS and NRPS systems, in different *Streptomyces* hosts using the PAC derivative pESAC13 (Table 3). The pESAC13 vector is a derivative of pPAC-S1 (Sosio *et al.* 2000). It contains an *oriT* locus for conjugal DNA transfer from *E. coli* and comes in two flavors: the original pESAC13, used for all work listed in Table 3, and the pESAC13A variant, carrying an additional selection marker for apramycin resistance (Fig. 3). The papers listed in Table 3 have enabled production of the desired metabolites in a genetically tractable host with sometimes yield improvements as compared to the native host or production of new congeners. It should be noted that all/most of the BGCs listed in Table 3 were isolated after screening large insert libraries in pESAC13 custom-produced by the specialized company Bio S&T (Montreal, Canada; www.biost.com).

Jankowitsch *et al.* (2012) expressed the 100-kb roseoflavin BGC from *S. davawensis* JCM 4913 (Fig. 2) in *S. coelicolor* M1152. The *S. tsukubaensis* NRRL 18488 BGC for FK506 (tacrolimus) (Fig. 2), a clinically important immunosuppressant, was introduced as part of a 130-kb ESAC clone into four different *S. coelicolor* strains (M512, M1146, M1152 and M1154). These strains all produced FK506 with significant yield improvements upon overexpression of a BGC-associated positive regulator (Jones *et al.* 2013). The utility of heterologous expression in *S. coelicolor* M1152 was also demonstrated for simocyclinone produced by *S. antibioticus* Tü6040 (Fig. 2). While the native host is particularly challenging for genetic manipulation, manipulation and functional analysis of the symocyclinone, BGC was possible in the heterologous system (Schäfer *et al.* 2015).

Upon screening an ESAC library of *S. toyocaensis* genomic DNA, Yim *et al.* (2016) identified a clone with a 140-kb insert containing the complete BGC for the glycopeptide A47934 (Fig. 2). After mobilizing it into *S. coelicolor* M1146, these authors were able to further decorate the A47934 scaffold following introduction of additional tailoring genes (Yim *et al.* 2016). Heterologous expression of azalomycin, a 36-membered aminopolyol macrolide, in *S. lividans* TK24 was observed using an ESAC clone carrying a 146-kbp insert (Hong *et al.* 2016). Production of azalomycin was achieved when the heterologous host was fed 4-guanidinobutyramide, a specific precursor of the azalomycin starter unit (Hong *et al.* 2016). Also, the heterologous host was able to methylate the guanidino group, notwithstanding that the cloned fragment did not contain the required methylase (Hong *et al.* 2016). The last reported example is related to the

Table 3. Examples of heterologous expression of BGCs cloned in BAC/PAC vectors.

Vector	Natural product (producer)	Biosynthetic class	BGC size (kbp)	Heterologous host	References
pStreptoBAC V	Daptomycin (*S. roseosporus* NRRL 11379)	NRP	128	*S. lividans* TK64 and TK23	Miao *et al.* 2005; Penn *et al.* 2006
	Iso-migrastatin (*S. platensis* NRRL18993)	PK-NRP	65	*S. lividans* K4-114, *S. coelicolor* M512, *S. albus* J1074, *S. avermitilis* SUKA4 and SUKA5	Feng *et al.* 2009
pSBAC	Meridamycin (*Streptomyces* sp. NRRL 30748)	PK-NRP	90	*S. lividans* K4-114[a]	Liu *et al.* 2009
	Tautomycetin (*Streptomyces* sp. CK4412	PK	80	*S. coelicolor* M145, *S. lividans* TK21	Nah *et al.* 2015
	Pikromycin (*S. venezuelae* ATCC 15439)	PK	60	*S. lividans* TK21, *S. coelicolor* M145	Pyeon *et al.* 2017
pESAC13	Roseoflavin (*S. davawensis* JCM 4913)	riboflavin (vitamin B2) analog.	106	*S. coelicolor* M1152	Jankowitsch *et al.* 2012; Schwarz *et al.* 2016
	FK506 (*S. tsukubaensis* NRRL 18488)	PK	83	*S. coelicolor* M512, M1146, M1152 and M1154	Jones *et al.* 2013
	Salinomycin (*S. albus* DSM 41398)	PK	120	*S. coelicolor* M1154	Luhavaya *et al.* 2014
	Anthracimycin (*Streptomyces* sp. T676)	PK	53	*S. coelicolor* M1146, M1152 and M1154	Alt and Wilkinson 2015
	Chaxamycin (*S. leeuwenhoekii*)	PK	80	*S. coelicolor* M1152	Castro *et al.* 2015
	Symocyclinone (*S. antibioticus* Tü6040)	PK	72	*S. coelicolor* M1152	Schäfer *et al.* 2015
	A47934 (*S. toyocaensis* NRRL 15009)	NRP	68	*S. coelicolor* M1146	Yim *et al.* 2016
	Azalomycin (*S. malaysiensis* DSM4137)	PK	99	*S. lividans* TK24	Hong *et al.* 2016
	Vancoresmycin (*Amycolatopsis* sp. *DEM30355*)	PK/NRP	141	*S. coelicolor* M1152	Kepplinger *et al.* 2017

PK = polyketide; NRP = non-ribosomalpeptide
[a]after cluster refactoring

identification, cloning and expression in *S. coelicolor* M1152 of vancoresmycin encoded by a 141-kbp PKS-NRPS BGC from *Amycolatopsis* sp. ST 101170 (Kepplinger *et al.* 2017) (Fig. 2). This represents so far the largest gene cluster to be expressed in a heterologous host and one of the few examples of successful intergeneric expression.

Recently, the broad-host range, transferable BAC vectors pSMART-BAC-S and pBAC-SBO have been developed by Lucigen (www.lucigen.com). These vectors have been used for the efficient preparation of two environmental libraries, highlighting their potential for discovering new compounds from the metagenome (https://www.lucigen.com/docs/posters/Soil-Metagenomics-Jan-13-SLAS.pdf).

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

A variety of ingenious cloning systems and vectors have been devised to harbor DNA fragments large enough to carry entire BGCs. Of the systems described above, the TAR, CATCH and IR systems require precise knowledge of the BGC(s) of interest. While conceptually simple, the IR approach requires an original producer that is genetically accessible, a requisite that limits this method to strains that have been well characterized. The TAR and CATCH systems make use of *in vivo* or *in vitro* recombination, respectively, and thus bypass the need to genetically access the native producer. These systems may be the best approach when the objective is mobilization of a single, large BGC. However, both systems are cluster-specific and cloning of multiple clusters from a single source would require working in parallel with different constructs. In contrast, BAC and PAC DNA libraries do not require *a priori* knowledge of the BGC(s) and can in principle provide the entire repertoire of BGCs from a single strain. However, the quality of the DNA library is essential, especially for large BGCs: ideally, the average insert size of the library should be larger than the target BGC, and the library should consist of about 10 genomic equivalents.

Potentially, these libraries can also provide a way for accessing BGCs from complex environments (i.e. metagenomics DNA libraries). Expression of environmental DNA for antibiotic discovery is an attractive proposition, but up to now only few BGCs, usually of the small size, have been expressed (Ongley *et al.* 2013; Katz, Hover and Brady 2016). Interestingly, BGCs have been built from smaller fragments using the TAR system, leading to the

assembly of 90-kb BGCs (Kim *et al.* 2010; Feng, Kallifidas and Brady 2011).

At the moment, the *E. coli-Streptomyces* shuttle BAC/PAC vectors seem the most suitable systems for cloning and expressing large BGCs. Using current recombination techniques, cloned clusters of interest can be retrofitted with desirable selectable markers, maintenance system or compound-modifying genes, thus adapting them to the specific needs of the investigator. The ease of generating whole genome sequences and the high-quality, large insert libraries that can be made in pESAC13 and similar vectors make these systems attractive for analyzing genetically intractable strains that are however (potential) producers of a large number of bioactive metabolites.

ACKNOWLEDGEMENT

We would like to thank Gianni Dehò for technical support and helpful discussion.

Conflict of interest. None declared.

REFERENCES

Alduina R, Gallo G. Artificial chromosomes to explore and to exploit biosynthetic capabilities of actinomycetes. *J Biomed Biotechnol* 2012;**2012**:1–10.

Alt S, Wilkinson B. Biosynthesis of the novel macrolide antibiotic anthracimycin. *ACS Chem Biol* 2015;**10**:2468–79.

Bierman M, Logan R, O'Brien K *et al.* Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* 1992;**116**:43–9.

Bilyk O, Sekurova ON, Zotchev SB *et al.* Cloning and heterologous expression of the grecocycline biosynthetic gene cluster. *PLoS One* 2016;**11**:e0158682.

Blin K, Kim HU, Medema MH *et al.* Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Brief Bioinform* 2017. doi: 10.1093/bib/bbx146. [Epub ahead of print].

Bonet B, Teufel R, Crüsemann M *et al.* Direct capture and heterologous expression of salinispora natural product genes for the biosynthesis of enterocin. *J Nat Prod* 2015;**78**:539–42.

Cano-Prieto C, García-Salcedo R, Sánchez-Hidalgo M *et al.* Genome mining of *Streptomyces* sp. Tü 6176: characterization of the nataxazole biosynthesis pathway. *ChemBioChem* 2015;**16**:1461–73.

Castro JF, Razmilic V, Gomez-Escribano JP *et al.* Identification and heterologous expression of the chaxamycin biosynthesis gene cluster from *Streptomyces leeuwenhoekii*. *Appl Environ Microb* 2015;**81**:5820–31.

Du D, Wang L, Tian Y *et al.* Genome engineering and direct cloning of antibiotic gene clusters via phage 8BT1 integrase-mediated site-specific recombination in *Streptomyces*. *Sci Rep* 2015;**5**:8740.

Feng Z, Kallifidas D, Brady SF. Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. *P Natl Acad Sci USA* 2011;**108**:12629–34.

Feng Z, Wang L, Rajski SR *et al.* Engineered production of isomigrastatin in heterologous *Streptomyces* hosts. *Bioorg Med Chem* 2009;**17**:2147–53.

Frengen E, Weichenhan D, Zhao B *et al.* A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites. *Genomics* 1999;**58**:250–3.

Fu J, Bian X, Hu S *et al.* Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat Biotechnol* 2012;**30**:440–6.

Galm U, Shen B. Expression of biosynthetic gene clusters in heterologous hosts for natural product production and combinatorial biosynthesis. *Expert Opin Drug Disc* 2006;**1**:409–37.

Genilloud O. Actinomycetes: still a source of novel antibiotics. *Nat Prod Rep* 2017;**34**:1203–32.

Gibson DG, Young L, Chuang RY *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 2009;**6**:343–5.

Gonzalez-Quiñonez N, López-García MT, Yagüe P *et al.* New ΦBT1 site-specific integrative vectors with neutral phenotype in *Streptomyces*. *Appl Microbiol Biot* 2016;**100**:2797–808.

Gregory MA, Till R, Smith MC. Integration site for *Streptomyces* Phage ΦBT1 and development of site-specific integrating vectors. *J Bacteriol* 2003;**185**:5320–3.

Hong H, Sun Y, Zhou Y *et al.* Evidence for an iterative module in chain elongation on the azalomycin polyketide synthase. *Beilstein J Org Chem* 2016;**12**:2164–72.

Ioannou PA, Amemiya CT, Garnes J *et al.* A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat Genet* 1994;**6**:84–89.

Jankowitsch F, Schwarz J, Rückert C *et al.* Genome sequence of the bacterium *Streptomyces davawensis* JCM 4913 and heterologous production of the unique antibiotic roseoflavin. *J Bacteriol.* 2012;**194**:6818–27.

Jiang W, Zhao X, Gabrieli T *et al.* Cas9-assisted targeting of chromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat Commun* 2015;**6**:9101.

Jiang W, Zhu TF. Targeted isolation and cloning of 100-kb microbial genomic sequences by Cas9-assisted targeting of chromosome segments. *Nat Protoc* 2016;**11**:960–75.

Jones AC, Gust B, Kulik A *et al.* Phage P1-derived artificial chromosomes facilitate heterologous expression of the FK506 gene cluster. *PLoS One* 2013;**8**:e69319.

Jordan PA, Moore BS. Biosynthetic pathway connects cryptic ribosomally synthesized posttranslationally modified peptide genes with pyrroloquinoline alkaloids. *Cell Chem Biol* 2016;**23**:1504–14.

Katz M, Hover BM, Brady SF. Culture-independent discovery of natural products from soil metagenomes. *J Ind Microbiol Biotechnol* 2016;**43**:129–41.

Kepplinger B, Morton-Laing S, Seistrup KH *et al.* Mode of action and heterologous expression of the natural product antibiotic vancoresmycin. *ACS Chem Biol* 2017; doi: 10.1021/acschembio.7b00733 [Epub ahead of print].

Kim JH, Feng Z, Bauer JD *et al.* Cloning large natural product gene clusters from the environment: Piecing environmental DNA gene clusters back together with TAR. *Biopolymers* 2010;**93**:833–44.

Li R, Xie Z, Tian Y *et al. polR*, a pathway-specific transcriptional regulatory gene, positively controls polyoxin biosynthesis in *Streptomyces cacaoi* subsp. asoensis. *Microbiology* 2009;**155**:1819–31.

Liu H, Jiang H, Haltli B *et al.* Rapid cloning and heterologous expression of the meridamycin biosynthetic gene cluster using a versatile *Escherichia coli-Streptomyces* artificial chromosome vector, pSBAC. *J Nat Prod* 2009;**72**:389–95.

Luhavaya H, Williams SR, Hong H *et al.* Site-specific modification of the anticancer and antituberculosis polyether salinomycin by biosynthetic engineering. *ChemBioChem* 2014;**15**:2081–5.

Martinez A, Kolvek SJ, Yip CL *et al*. Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl Environ Microb* 2004;**70**:2452–63.

Miao V, Coeffet-LeGal M, Brian P *et al*. Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology* 2005;**151**:1507–23.

Monaco AP, Larin Z. YACs, BACs, PACs and MACs: artificial chromosomes as research tools. *Trends Biotechnol* 1994;**12**: 280–6.

Monciardini P, Iorio M, Maffioli S *et al*. Discovering new bioactive molecules from microbial sources. *Microb Biotechnol* 2014;**7**:209–20.

Nah HJ, Pyeon HR, Kang SH *et al*. Cloning and heterologous expression of a large-sized natural product biosynthetic gene cluster in *Streptomyces* species. *Front Microbiol* 2017;**8**:394. eCollection 2017.

Nah HJ, Woo MW, Choi SS *et al*. Precise cloning and tandem integration of large polyketide biosynthetic gene cluster using *Streptomyces* artificial chromosome system. *Microb Cell Fact* 2015;**14**:140.

Ongley SE, Bian X, Neilan BA *et al*. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat Prod Rep* 2013;**30**:1121–38.

Penn J, Li X, Whiting A *et al*. Heterologous production of daptomycin in *Streptomyces lividans*. *J Ind Microbiol Biot* 2006;**33**:121–8.

Pyeon HR, Nah HJ, Kang SH *et al*. Heterologous expression of pikromycin biosynthetic gene cluster using *Streptomyces* artificial chromosome system. *Microb Cell Fact* 2017; **16**:96.

Rutledge PJ, Challis GL. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Micro* 2015;**13**:509–23.

Schäfer M, Le TB, Hearnshaw SJ *et al*. SimC7 is a novel NAD(P)H-dependent ketoreductase essential for the antibiotic activity of the DNA gyrase inhibitor simocyclinone. *J Mol Biol* 2015;**427**:2192–204.

Schwarz J, Konjik V, Jankowitsch F *et al*. Identification of the key enzyme of roseoflavin biosynthesis. *Angew Chem Int Ed* 2016;**55**:6103–6.

Shizuya H, Birren B, Kim UJ *et al*. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *P Natl Acad Sci USA* 1992;**89**:8794–7.

Simon R, Priefer U, Pühler A. A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in Gram negative bacteria. *Nat Biotechnol* 1983;**1**:784–91.

Sosio M, Bossi E, Donadio S. Assembly of large genomic segments in artificial chromosomes by homologous recombination in *Escherichia coli*. *Nucleic Acids Res* 2001 **29**:37e–37.

Sosio M, Giusino F, Cappellano C *et al*. Artificial chromosomes for antibiotic-producing actinomycetes. *Nat Biotechnol* 2000;**18**:343–5.

Stevens DC, Hari TP, Boddy CN. The role of transcription in heterologous expression of polyketides in bacterial hosts. *Nat Prod Rep* 2013;**30**:1391–411.

Tang X, Li J, Millán-Aguiñaga N *et al*. Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem Biol* 2015;**10**:2841–9.

Tracanna V, de Jong A, Medema MH *et al*. Mining prokaryotes for antimicrobial compounds: from diversity to function. *FEMS Microbiol Rev* 2017;**41**:417–29.

Yamanaka K, Reynolds KA, Kersten RD *et al*. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *P Natl Acad Sci USA* 2014;**111**:1957–62. Epub 2014 Jan 21.

Yim G, Wang W, Thaker MN *et al*. How to make a glycopeptide: a synthetic biology approach to expand antibiotic chemical diversity. *ACS Infect Dis* 2016;**2**:642–50.