J Ind Microbiol Biotechnol (2004) 31: 335–344
DOI 10.1007/s10295-004-0154-5

ORIGINAL PAPER

Igor A. Brikun · Andrew R. Reeves
William H. Cernota · Minh B. Luu · J. Mark Weber

The erythromycin biosynthetic gene cluster of *Aeromicrobium erythreum*

Received: 8 March 2004 / Accepted: 11 June 2004 / Published online: 15 July 2004
© Society for Industrial Microbiology 2004

Abstract The erythromycin-biosynthetic (*ery*) gene cluster of *Aeromicrobium erythreum* was cloned and characterized. The 55.4-kb cluster contains 25 *ery* genes. Homologues were found for each gene in the previously characterized *ery* gene cluster from *Saccharopolyspora erythraea*. In addition, four new predicted *ery* genes were identified. Two of the new predicted genes, coding for a phosphopantetheinyl transferase (*eryP*) and a type II thioesterase (*eryTII*), were internal to the *ery* cluster. The other two new genes, coding for a thymidine 5′-diphosphate-glucose synthase (*eryDI*) and a MarR-family transcriptional repressor (*ery*-ORF25), were found at the two ends of the *ery* cluster. A knockout in *eryDI* showed it to be essential for erythromycin biosynthesis. The gene order of the two *ery* clusters was conserved within a core region of 15 contiguous genes, with the exception of IS1136 which was not found in the *A. erythreum* cluster. Beyond the core region, gene shuffling had occurred between the two sides of the cluster. The flanking regions of the two *ery* clusters were not alike in the type of genes found.

Keywords *Aeromicrobium erythreum* · Erythromycin · Gene cluster · *eryDI* · *eryP* · *eryTI* · *eryTII*

Introduction

Research into erythromycin biosynthesis and genetics has focused nearly exclusively on the mycelial actinomycete, *Saccharopolyspora erythraea*, which is currently the only commercial producer of this macrolide antibiotic [2, 42]. The unicellular actinomycete, *Aeromicrobium erythreum*, however also produces erythromycin [13]. Previous studies with *A. erythreum* by Roberts et al. [31, 32] focused on the development of a transformation system and characterization of the erythromycin resistance gene *ermR* (formerly *ermA*). *A. erythreum* was assigned to the genus *Aeromicrobium* by Miller et al. [23], replacing its former classification in the genus *Arthrobacter* . Miller also developed a gene knockout system for this bacterium [22] using *Streptomyces*-based vectors. In addition, *Aeromicrobium erythreum* is reported to produce pure erythromycin A without accumulation of the erythromycin precursors B and C [13]. Recently, a study involving *A. erythreum* led to the development of a rational strain improvement strategy for erythromycin production, with general implications for the strain improvement of other natural products [30]. The effort by Reeves et al. [30] and this report show that *A. erythreum* is an attractive alternative system for the study of erythromycin production and has significant potential for commercial development.

The focus here was the cloning, characterization and mapping of the 25 genes of the erythromycin cluster, with respect to the already well characterized erythromycin gene cluster from *Saccharopolyspora erythraea*. The DNA sequence of a 61-kb region was determined, including the 55.4-kb *ery* cluster and flanking DNA on both sides. DNA sequencing was obtained by random transposon-mediated primer insertions into overlapping cosmid inserts, followed by extensive primer walking. The DNA sequencing experiments resulted in the creation of a single contiguous map of the region. Analysis of the DNA sequence resulted in the identification of at least four new *ery* genes.

Materials and methods

Bacterial strains, culture conditions and vectors

A. erythreum NRRL B3381 was obtained from the National Center for Agricultural Utilization Research

I. A. Brikun · A. R. Reeves · W. H. Cernota · J. M. Weber (✉)
Fermalogic Inc., 2201 W. Campbell Park Drive,
Chicago, IL 60612, USA
E-mail: mark.weber@fermalogic.com
Tel.: +1-312-7380050
Fax: +1-312-7380963

M. B. Luu
Rush University, 600 South Paulina Street, Suite 440,
Chicago, IL 60612, USA

(United States Department of Agriculture, Peoria, Ill.). Culture conditions for this organism were described by Roberts et al. [31]. Additional culture media were modified SCM [30] and 2xYT [30, 34]. The pBS cloning vector was used (Stratagene, La Jolla, Calif.) as was pFL8 [29]. Plasmids, cosmids and strains used in this study are summarized in Table 2.

Cloning of the *A. erythreum eryF* and *eryG* genes

A 1.15-kb *Age*I–*Kpn*I hybridization probe that spanned most of *eryF* of *S. erythraea* (accession number M54983) was labeled with digoxygenin (DIG) 11-dUTP according to the manufacturer's instructions (DIG Genius System; Roche Molecular Biochemicals, Indianapolis, Ind.) and hybridized to *A. erythreum* chromosomal DNA digested with *Bam*HI, revealing a 4-kb fragment. *Bam*HI-digested *A. erythreum* chromosomal DNA in the 4-kb to 5-kb range was excised from a 1% agarose gel and cloned into *Bam*HI-cut pBS (Stratagene). Following transformations into *Escherichia coli* DH5α, colony hybridizations were performed according to the DIG Genius system user's guide (Roche Molecular Biochemicals) to find clones containing the *A. erythreum eryF* gene. Plasmid inserts were sequenced showing that the *eryF* gene and neighboring genes, including the entire *eryG* gene, were cloned. The pBS *eryF* clones were double-digested with *Eco*RI–*Hin*dIII, ligated to similarly digested pFL8 [29] and transformed into *E. coli* DH5α. The new pFL8 derivative carrying the clone from pBS *eryF* was designated pFL *eryF*.

Preparation of the *A. erythreum* cosmid library

Total cellular DNA was isolated from 100-ml cultures of *A. erythreum* B3381 cells grown in modified SCM medium. Cells were embedded into 0.7% agarose beads and then treated with protease to expose the DNA for further processing. A partial restriction endonuclease digest of the DNA by *Mbo*I (Gibco BRL, Rockville, Md.) was done by varying the concentration of the enzyme. Fragments approximately 40 kb in length were separated by pulsed-field gel electrophoresis in 1% low melting agarose with an initial switch time of 2.5 s and a final switch time of 3 s during a 16-h run at 6 V/cm (Chef DRIII system; BioRad, Hercules, Calif.). DNA fragments were purified by agarase (Roche Molecular Biochemicals) treatment according to the supplier's protocol. Purified 40-kb DNA fragments were ligated to SuperCos cosmid vector (Stratagene) and packaged into phage particles using Gigapack III gold packaging extract (Stratagene). Titration and propagation of the packaged extracts were done in *E. coli* strain XL1BlueMR. Approximately 1,200 clones (estimated to be six genomic equivalents) were propagated in 96-well microtiter plates, each well containing 100 μl of 2xYT medium supplemented with 40 μg/ml of kanamycin. For storage at −80°C, the culture broth and cells were supplemented with glycerol to a final concentration of 20%.

Screening the *A. erythreum* cosmid library to find the erythromycin gene cluster

Cosmids containing the *ery* gene cluster were found using the rows-and-columns method [16]. Plasmid DNA was extracted using either the Concert high purity plasmid miniprep system (Gibco BRL) or the Perfectprep plasmid mini kit (Eppendorf, Westbury, N.Y.). The DNA preparations were then screened by PCR using the *A. erythreum eryG* gene primers (ATGGCAGCACGACAGAGCATCGTG, CTGACGCTCGACTTGAAGCTG) and the *A. erythreum eryAI* gene primers (TGCCACTTCTAGTTTGTCCAG, GGCGACGGCGTCGAG). Screening was performed by PCR in 96-well microtiter plates using a PTC-100 amplifier (MJ Research, Waltham, Mass.), with a pre-heating step at 96°C for 2 min, followed by 34 cycles of 96°C for 1 min, 56°C for 30 s and 72°C for 1 min.

DNA sequencing and analysis of *A. erythreum* cosmid DNA

Sequencing reactions were performed using the Big Dye terminator cycle sequencing kit (Applied Biosystems, Foster City, Calif.) according to the manufacturer's protocol. DNA prepared by PCR was precipitated with ethanol; and DNA sequencing was performed by Qiagen sequencing services (Bothell, Wash.). Editing and assembly of the sequences was done using SEQUENCHER software (Gene Codes Corp., Ann Arbor, Mich.). All sequences were analyzed by the NCBI BLASTX or BLASTP programs with default settings.

The transposon EZ::TN <TET-1> (Epicentre, Madison, Wis.) was used for the in vitro transposition reactions of selected cosmid clones. The first 24 transposon-containing clones, derivatives of 7F9, 7D12 and 2B9 (Fig. 1), were analyzed by *Eco*RI restriction endonuclease digestion with regards to the randomness of the transposition event and the frequency of transposon integration into the vector region of the cosmid. This analysis revealed that all selected transposon-mutagenized clones contained the transposon in different sites and only one contained a transposon in the vector.

A total of 135 transposon insert-generated sequencing reactions were analyzed in cosmid 7F9, 68 reactions were generated with cosmid 7D12 and 66 reactions were generated with cosmid 2B9. Approximately 30 contigs were assembled from transposon-based sequencing; and 14 of the contigs contained genes that belonged to the *ery* gene cluster. The total length of these contigs was approximately 40 kb.

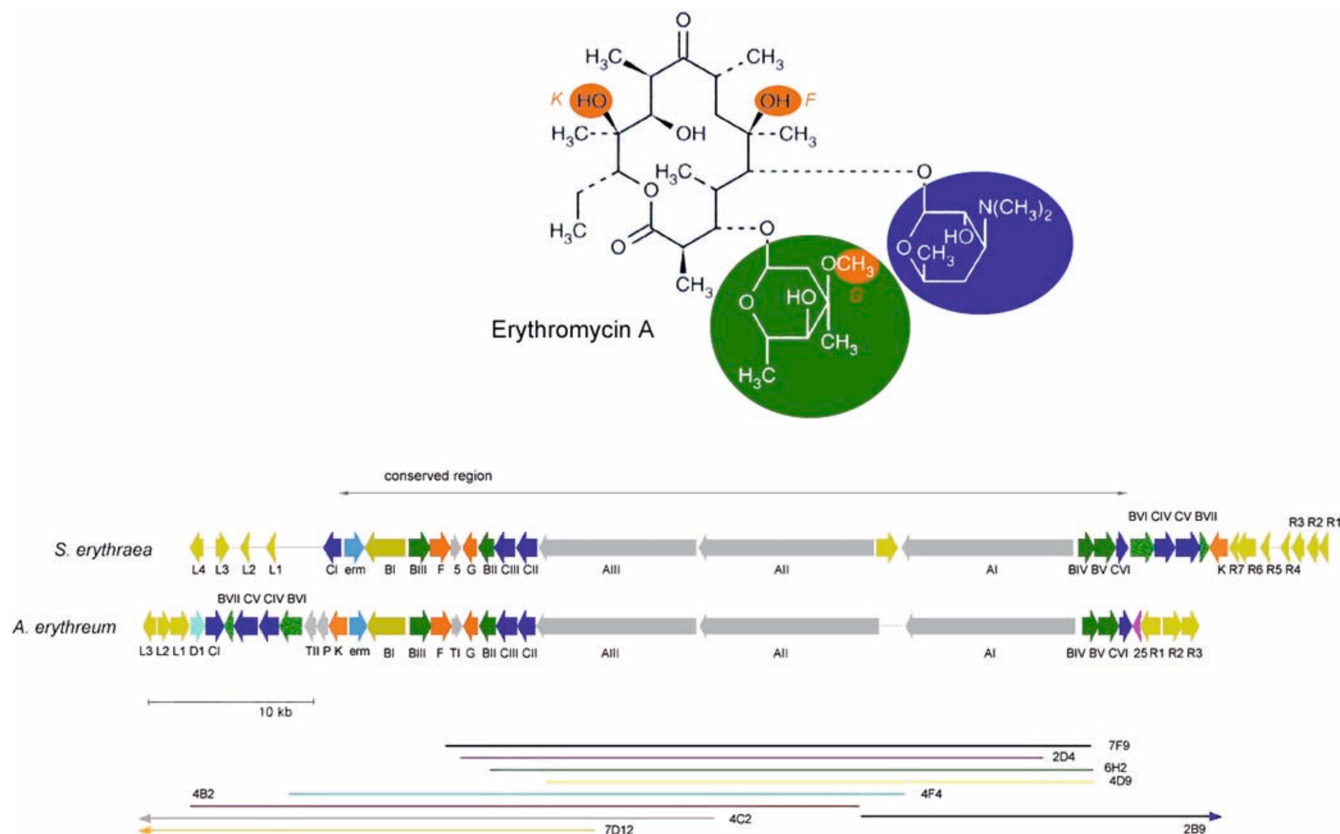

Fig. 1 A comparison of erythromycin gene clusters from *A. erythreum* and *S. erythraea*. The erythromycin A molecule (*top*) is above the two gene maps, with colors correlating the components of the structure to the genes below. Gene designations are abbreviated: e.g., *L4* ORFL4, *CI eryCI* (see Table 1 for complete list). The polyketide portion of the molecule is not colored and the corresponding *eryA* genes are *gray arrows*. Flanking ORFs on both sides are shown in *yellow*. *eryBI* (*ery*-ORF2) is not in green because *eryBI* mutations in *S. erythraea* did not confer the EryB phenotype [15]. Both clusters have copies of *eryBI* and more reason now exists to assume this gene plays some role in erythromycin biosynthesis. IS1136 (*IS*) of *S. erythraea* is shown in *yellow* since it is not involved in erythromycin biosynthesis. In *A. erythreum*, the lack of IS1136 is indicated by the *dotted line* between *eryAI* and *eryAII*. The erythromycin resistance genes, *erm* in *light blue*, indicate the two homologues *ermE* (*S. erythraea*) and *ermR* (*A. erythreum*). The tailoring genes *eryF*, *eryG* and *eryK* (and their corresponding functional groups), are shown in *orange*. The MarR-family transcriptional regulator, *ery*-ORF25, is shown in *pink*. Cosmid inserts isolated from *A. erythreum* that overlap the *ery* gene cluster are shown at the *bottom*. *Arrows* indicate where cosmids continue beyond the *ery* cluster region. Cosmids are labeled with their alphanumeric designation (see Table 2)

Primer-walking was used to fill the gaps between the contigs. In the first step, 17 pairs of primers were designed from both ends of each contig and used to generate the sequences that extended the length of the contigs by approximately 500 bp from each side. Extended contigs were assembled and the next set of primers was synthesized. The primer-walking step was repeated four times. In the second step, 11 pairs of primers were used, five pairs in the third step and four pairs in the fourth step. As a result of the primer-directed sequencing, all erythromycin-biosynthetic genes (with adjacent regions) were assembled into one contig.

Protoplast generation and transformation

A. erythreum protoplasts were generated and transformed using the method described by Reeves et al. [30].

Plasmid eviction procedure

An example for the knockout of *eryG* is presented. Primary transformants of *A. erythreum* containing integrated pFL2088 (Table 2) by single-crossover insertion, were thiostrepton- and kanamycin-resistant. The desired gene replacement or double-crossover strains (evictants) that lost the plasmid backbone and retained the *eryG* insertion mutation in the chromosome, would be thiostrepton-sensitive and kanamycin-resistant. In order to obtain the double-crossover derivatives, single-crossover strains were patched onto 2xYT[G] agar [30, 34] containing kanamycin (10 μg/ml) and then streaked for single colonies on fresh media containing kanamycin. After sufficient growth, two single colonies were transferred to separate 4-ml 2xYT[G] broth test-tube cultures containing kanamycin (10 μg/ml) and incubated at 32°C and 350 rpm for 5 days. On day 5, cells from each of the test tube culture were streaked for single colonies onto 2xYT[G]

Table 1 Deduced function of genes in the erythromycin gene cluster and flanking sequences in *A. erythreum*. *NK* Not known, *NA* not available

ORF	*A. erythreum* gene[a]	*S. erythraea* gene[b]	Proposed function of *A. erythreum* gene[c]	Identity (%):size (number of amino acids)[d] (%)	Sequential accession number[e]	Reference(s)
L3	ORFL3 (fragment)	NK	Transcriptional repressor, TetR family	53:25	AAN69554.1	[12, 25]
L2	ORFL2	NK	Dehydrogenase/reductase, *fabG* family	30:251	AAC62538.1	[4]
L1	ORFL1	NK	Unknown	NA	NA	NA
1	*eryD1*	NK	dTDP-glucose synthase	65:288	AAN28687.1	[20, 33, 44]
2	*eryC1*	*eryCI*	Transaminase	59:365	CAA33548.1	[7, 33, 43, 46]
3	*eryBVII*	*eryBVII*	Deoxyhexose epimerase	66:193	AAB84077.1	[14, 15, 33, 38]
4	*eryCV*	*eryCV*	Deoxyhexose reductase	70:489	CAA72085	[14, 33, 38]
5	*eryCIV*	*eryCIV*	Deoxyhexose dehydratase	58:401	AAB84075.1	[14, 33, 38]
6	*eryBVI*	*eryBVI*	Deoxyhexose dehydratase	53:510	CAA72083.1	[14, 33, 38]
7	*eryTII*	NK	Thioesterase, type II	37:247	AAA26497.2	[18, 19, 47]
8	*eryP*	[SePptII][d]	Phosphopantetheinyl transferase	45:220	AAR92400.1	[48]
9	*eryK*	*eryK*	Erythromycin C-12 hydroxylase	73:397	P48635	[37]
10	*ermR*	*ermE*	Ribosomal RNA methyltransferase	58:381	CAB60001.1	[8, 31, 39, 40]
11	*eryBI*	*ery*-ORF2 (*eryBI*)	*β*-Glucosidase	71:808	CAA74702.1	[14, 15, 46]
12	*eryBIII*	*eryBIII*	Methyltransferase	73:414	CAA74703	[15, 18, 47]
13	*eryF*	*eryF*	Cytochrome P450 oxidoreductase	70:403	AAA26496	[18, 47]
14	*eryTI*	*ery*-ORF5	Thioesterase	56:247	AAA26497.2	[18, 47]
15	*eryG*	*eryG*	Methylase	72:306	CAA42929	[18, 26, 45, 46, 47]
16	*eryBII*	*eryBII*	Deoxyhexose reductase	72:333	AAB84068.1	[18, 33, 38, 46]
17	*eryCIII*	*eryCIII*	L-Desosaminyltransferase	69:421	CAA74710.1	[18, 33]
18	*eryCII*	*eryCII*	Deoxyhexose isomerase	41:361	CAA74711.1	[18, 33, 46]
19	*eryAIII*	*eryAIII*	Polyketidesynthase modules 5 and 6	30–58%	AAA26495.1	[6, 9, 46]
20	*eryAII*	*eryAII*	Polyketidesynthase modules 3 and 4	45%	AAA26494.1	[1, 9]
	none	IS1136	Insertion sequence	NA	AAA26505.2	[10]
21	*eryAI*	*eryAI*	Polyketidesynthase modules 1 and 2	46–63%	AAA26493.2	[9]
22	*eryBIV*	*eryBIV*	Deoxyhexose dehydratase	49:322	AAB84071.1	[14, 38]
23	*eryBV*	*eryBV*	L-Mycarosyltransferase	55:415	AAB84072.1	[14, 38]
24	*eryCVI*	*eryCVI*	*N*-Methyltransferase	68:237	AAB84073.1	[14, 38]
25	*ery*-ORF25	NK	*marR* family transcriptional regulator	45:121	CAB48917.1	[17]
R1	ORFR1	NK	Putative oxydoreductase	43:1151	CAB53292	NA
R2	ORFR2	NK	*lipN* (*Mycobacterium tuberculosis*)	42:337	NP_217486.1	[41]
R3	ORFR3 (fragment)	NK	Monoamine oxidase	47:311	P49253	[5]

[a]Erythromycin gene cluster from *A. erythreum*
[b]Erythromycin gene cluster from *S. erythraea*
[c]Probable function based on BLASTP results
[d]Percent identity from BLASTP results, a colon, size of the region of homology as a number of contiguous amino acids[e]Accession numbers were selected from the erythromycin gene cluster of *S. erythraea*, if one existed. If more than one accession number existed, only one is reported. If no homologue from *S. erythraea* existed, a representative homologue is given from another organism, with reference if available

Table 2 Plasmids, cosmids, strains and transposons used in this study. Square brackets denote antibiotic resistance in both *E. coli* and *A. erythreum*. Curved brackets denote antibiotic resistance in *A. erythreum* only. No brackets denote antibiotic resistance in *E. coli* only

Item	Description	References
Plasmids, *E. coli*		
SuperCos	Cosmid vector used for generating the *A. erythreum* genomic library in *E. coli*. Ap^r Kn^r	Stratagene
pFL8	A derivative of pBS(+) containing the *Streptomyces* origin of replication from pIJ487. Used for general cloning and integration of homologous DNA fragments into the *S. erythraea* chromosome. Ap^r ($Thio^r$)	[29]
pFL1324	A derivative of pBS(+) containing a 3,869-bp *Bam*HI *A. erythreum* chromosomal fragment spanning the *eryF* region. Contains all of the *eryF*, *eryG* and *eryTII* genes and part of the *eryBII* and *eryBIII* genes. Ap^r	This study
pFL1375	A derivative of pFL8 containing a 3,869-bp *Bam*HI *A. erythreum* chromosomal fragment spanning the *eryF* region. Contains all of the *eryF*, *eryG* and *eryTII* genes and part of the *eryBII* and *eryBIII* genes. Ap^r ($Thio^r$)	
pFL2072	A derivative of pUC19 containing the 3,869-bp *Bam*HI fragment from pFL1375. Ap^r	
pFL2078	A derivative of pFL2072 containing a kanamycin resistance gene cassette cloned into the *SphI* site in *eryG*. Used as an intermediate plasmid for disrupting the *A. erythreum eryG* gene. Ap^r [Kn^r]	
pFL2082	Vector containing the *aph1* gene from pUC4 K (Amersham–Pharmacia, Piscataway, N.J.) cloned into the *Ssp*I site of pUC19. Used as an integration vector for making insertions into the *A. erythreum* chromosome. Ap^r [Kn^r]	[30]
pFL2087	Plasmid used to obtain the thiostrepton resistance gene on a 1.041-kb *Kpn*I/*Xmn*I fragment. Ap^r Tet^r ($Thio^r$)	[30]
pFL2088	A derivative of pFL2078 containing the thiostepton resistance gene from pFL2087 cloned into the *Kpn*I and *Xmn*I sites. Used for knocking-out the *A. erythreum eryG* gene. Ap^r [Kn^r] ($Thio^r$)	This study
pFL3085	A derivative of pCRII containing a 770-bp amplified fragment of the *ery*-ORF25 gene and flanking regions. Intermediate plasmid for generating a knockout of *ery*-ORF25. Ap^r Kn^r	
pFL3086	A derivative of pFL3085 containing a thiostrepton resistance gene cloned into a unique *Sma*I site in *ery*-ORF25. Ap^r [Kn^r] ($Thio^r$)	
Cosmids, *E. coli*		
7F9	SuperCos containing the *A. erythreum eryAI*, *eryAII* and *eryAIII* genes along with the left flanking region spanning from *eryCII* to *eryBIII*. Used for sequencing all of the *eryA* genes and part of the left flanking region. Ap^r Kn^r	This study
2B9	SuperCos containing part of the *A. erythreum eryAIII* gene and all of the *eryAI* and *eryAII* genes along with the right flanking region covering *eryBIV* to ORFR3. Used for sequencing *eryAI*, *eryAII* and the right flanking region. Ap^r Kn^r	
7D12	SuperCos containing part of the *A. erythreum eryAIII* gene along with the entire left flanking region. Used for sequencing the left flanking region. Ap^r Kn^r	
2D4, 6H2, 4D9, 4F4, 4B2, 4C2	Additional cosmids from the *A. erythreum* total DNA library that overlap the *ery* cluster but were not needed for DNA sequencing (see Fig. 1)	

Table 2 (Contd.)

Item	Description	References
Strains, *A. erythreum*		
FL262	NRRLB-3381. Wild-type strain; erythromycin producer	[13]
FL2105	Double-crossover knockout of *eryG* derived by integration and eviction of pFL2088. Thio^s Kn^r	This study
FL2223	*ery*-ORF25 knockout strain derived by double-crossover insertion of the thiostrepton resistance gene using pFL3086. Thio^r Kn^s	
FL3055	Double-crossover knockout of *eryD1* derived by integration and eviction of pFL3055. Thio^r Kn^r	
Strain, *E. coli*		
DH5α-e	Host strain for electroporation to create *A. erythreum* genomic library	Invitrogen
Transposon		
EZ::TN <TET-1>	Tet^r Tn5 derivative used for in vitro mutagenesis and sequencing	Epicentre, Madison, Wis.

agar containing kanamycin and incubated at 32°C. After 5 days of growth, 100 single colonies were tested for their antibiotic phenotypes by patching onto 2xTY[G] plates containing kanamycin, thiostrepton or no antibiotic addition. Of the 100 colonies patched, 98 isolates were gene replacement strains (kanamycin-resistant, thiostrepton-sensitive). These strains were designated FL2105.

Screening the *S. erythraea* genome for the *A. erythreum eryTII* gene

To screen for a type II thioesterase (*eryTII*) gene in *S. erythraea*, PCR primers were designed based on highly conserved regions in *eryTII*. The primer sequences were 5′-CGGAATTCCCGCACGCCGGCGGGTCTGCC-3′ and 5′-TTGAAGTTCGGCCACGACCGCGCCCATG-CTGTGACCGAA-3′. Each primer had a restriction endonuclease site engineered at the 5′ end to facilitate later cloning steps. The PCR program used for amplification began with a denaturation step at 95°C for 2 min, followed by 25 cycles of a heating step at 95°C for 1 min, an annealing step at 55°C and an extension step at 72°C. A final extension step at 72°C was used to finish off any sequence gaps. Taq DNA polymerase 2× PCR mix (Fermentas, Vilnius, Lithuania) was used with 2 mM Mg^{2+} . After amplification, the product was excised from a 1% agarose gel and ligated into the pGEM-T Easy vector (Promega, Madison, Wis.) and its DNA sequence was determined.

Construction of pFL2088 for the *eryG* knockout

The 3.9-kb *Bam*HI insert of pFL *eryF* was subcloned into *Bam*HI-digested pUC19 and designated pFL2072. To disrupt *eryG* in pFL2072 the plasmid was first partially digested with *Sph*I. pFL2072 contains two *Sph*I sites, one in the *eryG* gene and the other in the polylinker. *Sph*I-digested pFL2072 was used in a ligation reaction with the kanamycin-resistance gene, *aph1*, from pUC4K (Pharmacia Biochemicals, Piscataway, N.J.) and engineered to carry *Sph*I ends. The ligation mix was transformed into *E. coli* and plated on LB agar containing ampicillin and kanamycin, both at 50 μg/ml. Plasmids from kanamycin-resistant transformants were analyzed by restriction endonuclease analysis; and those showing evidence of an *eryG*-insertion knockout were designated pFL2078.

The final cloning step involved inserting the thiostrepton resistance gene into the plasmid backbone of pFL2078. A 1-kb *Xmn*I–*Kpn*I fragment from pFL2087 was used as the source of the thiostrepton-resistance gene. Plasmids were determined to contain the correct thiostrepton-resistance gene insert after restriction endonuclease analysis with *Kpn*I and *Xmn*I in double digestions and *Pvu*II digestion. This plasmid was designated pFL2088.

Knockouts of the *eryG* gene were obtained after transformation of *A. erythreum* protoplasts [30] with pFL2088 and selection of primary transformants with thiostrepton [10 μg/ml], followed by plasmid eviction and screening for thiostrepton-sensitive and kanamycin-resistant plasmid evictants (see above).

Construction of pFL3086 for the *ery*-ORF25 knockout

A DNA fragment (770 bp) corresponding to *ery*-ORF25 and a short part of the flanking genes was amplified from *A. erythreum* chromosomal DNA, using primers CACGCTGTTCGAGCGCGAG and GTGGTCAA-

CACGATGTTCCG. The amplified DNA fragment was cloned into pCRII (Invitrogen, Carlsbad, Calif.), resulting in plasmid pFL3085. The thiostrepton-resistance gene, with *Sma*I and *Dra*I blunt ends, was inserted into the *Sma*I site of *ery*-ORF25 in plasmid pFL3085.

The resulting plasmid, pFL3086, was transformed into *A. erythreum* [30] to knock out the *ery*-ORF25 gene by a double-crossover (gene replacement) event, inserting the thiostrepton resistance gene into *ery*-ORF25 in the chromosome to create strain FL2223. Strain FL2223 was obtained from the direct screening of primary transformants for thiostrepton resistance and kanamycin sensitivity, and did not require the plasmid eviction procedure.

Construction of pFL3055 for the *eryDI* knockout

Two 2-kb DNA fragments flanking the thymidine 5′-diphosphate (dTDP)-glucose synthase *eryDI* produced by PCR, using primers described below, were cloned into plasmid pFL2082 along with the thiostrepton-resistance gene, which was inserted between them in a four-component ligation.

Each DNA fragment used in the four-component ligation, besides the plasmid component, was initially cloned into pGEM-T Easy. The two 2-kb DNA fragments flanking *eryDI* and the thiostrepton-resistance gene were cut out of the pGEM-T derivatives, purified from agarose gels and then used in the four-component ligation with pFL2082 to form pFL3055. pFL3055 was transformed into *A. erythreum* protoplasts and *eryDI* knockouts were selected by thiostrepton resistance, screened for kanamycin-sensitivity and then verified by PCR and DNA sequencing.

The primers used to generate the left-side 2-kb DNA fragment were ggaattcCGCGCATGACTCGACCCTAG and cgggatccGAGTGACCATGTCCACGCAC; and the primers for the right-side 2-kb fragment were cccaagcttCTACCGCGTCTGAGCAGGC and acatgcatgcGACCTCGACGGAGCCCAC. The primers for amplification of the thiostrepton-resistance gene were ggaattcTCTGATTAATAAGATGATCCCCGG and cccaagcttGCTTATCGGTTGGCCGCGAGA.

Results and discussion

New *ery* genes identified from *A. erythreum*

Four *ery*-ORFs appear in the *ery* cluster from *A. erythreum* (accession number AY623658) that do not appear in the previously characterized *ery* cluster of *S. erythraea*. The *ery*-ORF1 predicted gene (*eryDI*), occurs at the left end of the cluster (Table 1, Fig. 1) and is a presumptive dTDP-glucose synthase. This enzyme converts D-glucose-1P to dTDP-glucose, a step common to the biosynthesis of both sugars in the erythromycin molecule.

A plasmid insertion knockout in *eryDI* created a block in erythromycin biosynthesis, resulting in the accumulation of erythronolide B (data not shown). Results of cross-feeding experiments for this mutant were consistent with the EryD phenotype described previously for the *eryD24* mutation [44], in which EryD is the designation for blocks affecting early sugar biosynthesis. The *ery*-ORF1 predicted gene was assigned the name *eryDI*, because it was the first gene of the EryD phenotype to be identified. It is not known whether the *S. erythraea eryD24* mutation [44] is in the homologue of the *A. erythreum eryDI* gene.

The *ery*-ORF7 gene (*eryTII*) is predicted to code for a type II thioesterase. A second type II thioesterase (*eryTI*) is also located within the *A. erythreum ery* cluster: it is the homologue to the *ery*-ORF5 gene in the *S. erythraea ery* cluster, between the *eryF* and *eryG* genes (Fig. 1).

In an experiment to determine whether *S. erythraea* has a homologue to the *eryTII* gene of *A. erythreum*, the *S. erythraea* genome was probed by PCR using primers designed from the *eryTII* of *A. erythreum* (see Materials and methods). A PCR product of the expected size (225 bp) was produced and cloned and its DNA sequence was determined. The DNA sequence of the cloned fragment indicated it to be a homologue to the *eryTII* gene of *A. erythreum*. Thioesterases are reported to play a role as editing enzymes, removing aberrant starter units attached to the acyl carrier domain of the polyketide synthase or non-ribosomal peptide synthase [3, 19, 36, 49]. The abundance of this protein in the cell has an effect on the antibiotic production capacity of the cell. Without this gene in some systems, production of the antibiotic is significantly reduced.

The *ery*-ORF8 gene is predicted to code for a phosphopantetheinyl transferase (PPTase). PPTases are normally required for fatty acid, polyketide and non-ribosomal peptide biosynthesis. They activate the synthetases through a post-translational modification of acyl carrier proteins (ACPs) in which a phosphopantetheinyl group is attached to a conserved serine residue of the ACP. PPTases fall into three different families, based on their substrate specificity and their involvement in primary or secondary metabolism [11, 24]. The *ery*-ORF8 gene shows the highest homology to the promiscuous PPTases in the *Sfp* and *Svp* families [35]. No PCR products were found when the *S. erythraea* genome was probed by PCR, using primers designed from the *ery*-ORF8 DNA sequence of *A. erythreum* (see Materials and methods). Recent genomic sequencing of *S. erythraea*, however, uncovered a gene coding for a discrete PPTase; and evidence from experiments performed in *E. coli* indicated it could play a role in erythromycin biosynthesis in *S. erythraea* [48]. The PPTase from *S. erythraea* has 44% identity to *ery*-ORF8 from the *A. erythreum ery* gene cluster. Because of its location in the *ery* cluster and probable role in

erythromycin biosynthesis, it was designated a PPTase (*eryP*).

The predicted gene, *ery*-ORF25, lies at the far right end of the *ery* cluster (Fig. 1). BLASTP analysis revealed its gene product has similarity to transcriptional repressors in the MarR family. MarR is a protein dimer, containing a double-winged helix DNA-binding motif [17]. Mar (multiple antibiotic resistance) protein expression is controlled by the MarR repressor in *E. coli*. Since *ery*-ORF25 is immediately adjacent to known genes of the *ery* cluster and because of its presumed role in regulation of antibiotic resistance, it is tentatively included as a part of the *ery* gene cluster. However, a plasmid insertion knockout of *ery*-ORF25 (see Materials and methods) had no noticeable effect on erythromycin biosynthesis when the mutant was grown in modified SCM media.

Organization of the *A. erythreum ery* cluster

The organization of genes in the *ery* cluster from *A. erythreum* is described in relation to the previously characterized *ery* cluster of *S. erythraea*. The 55.4-kb *A. erythreum ery* cluster contains 25 *ery*-ORFs, four of which lack homologues in the 52.5-kb *S. erythraea ery* gene cluster (Fig. 1, Table 1). A homologue of each gene in the *S. erythraea ery* cluster can be found in the *A. erythreum ery* cluster, except for the IS element (IS1136), located between *eryAI* and *eryAII* in *S. erythraea* [10].

A group of 15 genes from *ermR* to *eryCVI* showed an identical gene order and a high degree of homology to the *S. erythraea ery* cluster. The *eryA* genes for the polyketide synthase show the same internal organization of modules and domains as the three *eryA* genes of *S. erythraea* . The absence of the IS element between *eryAI* and *eryAII* in *A. erythreum* is the only notable difference in this region.

Although the overall order and arrangement of genes in the central core region is nearly identical to *S. erythraea*, one notable difference lies in the arrangement of the convergent pair of genes, *eryG* and *ery-ORF5* . In *S. erythraea*, *eryG* and *ery*-ORF5 overlap by 28 bp in their 3′ regions [47]. The corresponding pair of genes in *A. erythreum* (*eryG* and *eryTI*) do not overlap (Fig. 2). Studies in other organisms have shown that convergent overlapping genes suffer from impaired expression, due in part to the binding of the transcripts in a sense–antisense arrangement. The *eryG* gene of *A. erythreum* is nearly identical in size and amino acid sequence to the *eryG* from *S. erythraea* (72% over the entire length of the protein) and is not a truncated derivative of *eryG*. It is not known whether the *eryG*–*ery*-ORF5 overlap has a negative effect on the expression of *eryG* or on the efficiency of the conversion of erythromycin C to erythromycin A.

The second subcluster of genes, whose gene order is conserved between both organisms, is the *eryBVII*, *CV*, *CIV* and *BVI* subcluster. All four of these genes are involved in sugar biosynthesis, two are mycarose-related and two are desosamine-related. They all share approximately the same high degree of homology to the *S. erythraea* genes, but their gene order is inverted and their location is on the opposite side of the central conserved region.

The *eryCI* and *eryK* genes were shuffled independently of other *ery* genes. The *eryK* gene of *A. erythreum* occupies the place of the *eryCI* gene in *S. erythraea*, upstream and divergently transcribed from the erythromycin resistance gene, *ermR*.

The *ery* cluster flanking regions

The left flanking region begins with ORFL3 (Fig. 1), a partial ORF, identified through BLASTX analysis as a homologue of transcriptional regulatory proteins in the TetR family [12].

ORFL2 codes for a predicted protein showing high homology to several families of short-chain dehydrogenases/reductases. The closest match was to the FabG family of dehydrogenases. FabG, from *E. coli*, encodes a 3-ketoacyl-ACP reductase and is required for fatty acid biosynthesis [28]. In *Pseudomonas aeruginosa*, two homologues of FabG have been characterized. One FabG homologue, in the *fab* (fatty acid biosynthesis) gene cluster, is involved in primary metabolism [21] and the

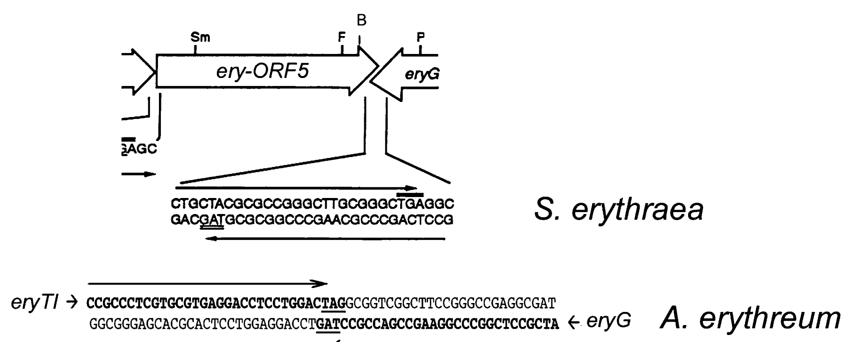

Fig. 2 DNA sequence comparison of the *eryG*–*ery*-ORF5 overlapping genes in *S. erythraea* with the equivalent region of *A. erythreum*, where no overlap is found. The *ery*-ORF5 gene of *S. erythraea* and *eryT1* of *A. erythreum* are homologous type II thioesterases. *Arrows* above the DNA sequence data indicate the predicted direction and extent of transcription of the genes shown. *Sm*, *Sma*I, *F*, *Fok*I, *P*, *Pst*I

second homologue, *rhlG*, is involved in secondary metabolism (rhamnolipid biosynthesis) and is found elsewhere in the chromosome [4].

BLASTX analysis of ORFL1 did not reveal strong homologies to known genes in the public databases. ORFL1 has a clear ribosome-binding site and start codon separated by 6 bp and also has a stop codon appearing just before the beginning of *ery*-ORF1 (*eryDI*).

All three of the left-side flanking ORFs could potentially be involved in erythromycin biosynthesis. However, additional experiments will be necessary to determine their role, if any, in this process.

The right flanking region contains three ORFs adjacent to *ery*-ORF25. These genes and their predicted function are listed in Table 1. ORFR1 and ORFR2 have homologies to genes involved in lipid metabolism. These genes could play a role in oil utilization; and soybean oil is an important component of the erythromycin production medium. ORFR3 is a predicted monoamine oxidase. Again, mutational analyses of the ORFs in the right-side flanking region will be helpful to determine whether any of these genes plays a direct or indirect role in erythromycin biosynthesis.

Comparisons of the flanking regions of *S. erythraea* [27, 29] and *A. erythreum ery* clusters reveal no obvious similarities to one another.

Targeted disruption of *eryG*

Disruption of the *eryG* gene in *S. erythraea* results in the accumulation of erythromycin C [45]. When an *eryG* knockout was performed in *A. erythreum*, using plasmid pFL2088 (see Materials and methods), the same result was obtained.

Conclusions

Analysis of the *A. erythreum ery* gene cluster (accession number AY623658) revealed four new predicted *ery*-ORFs: *eryDI*, *eryTII*, *eryP* and a predicted regulatory gene in the *marR* family (*ery*-ORF25).

The *eryDI* gene of *A. erythreum* is the first gene of the EryD class to be cloned, sequenced and mutationally characterized. The *eryG* knockout of *A. erythreum* confirmed its involvement in the production of erythromycin A in this organism. The two *ery* clusters show a high degree of conservation in gene order, especially in a central core region of 15 genes from the *erm* genes to *eryCVI*. The flanking genes of the two clusters show no conservation of gene identity or order. Mutational analyses of these genes could lead to additional *ery* genes being found. The *eryG* and *eryTI* genes of *A. erythreum* do not overlap, as they do in *S. erythraea*; and *S. erythraea* has a homologue to the *eryTII* gene of *A. erythreum*.

Acknowledgements The authors thank Eric Miller and Roy Wesley for helpful discussions and Roy Wesley for technical assistance. The National Institute of General Medical Sciences (National Institutes of Health, Small Business Innovation Research award R43GM63278) provided the financial support for this study.

References

1. Bevitt DJ, Cortes J, Haydock SF, Leadlay PF (1992) 6-Deoxyerythronolide-B synthase 2 from *Saccharopolyspora erythraea*. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. Eur J Biochem 204:39–49
2. Bunch RL, McGuire JM (1953) Erythromycin, its salts, and method of preparation. US patent 2,653,899
3. Butler AR, Bate N, Cundliffe E (1999) Impact of thioesterase activity on tylosin biosynthesis in *Streptomyces fradiae*. Chem Biol 6:287–292
4. Campos-Garcia J, Caro AD, Najera R, Miller-Maier RM, Al-Tahhan RA, Soberon-Chavez G (1998) The *Pseudomonas aeruginosa rhlG* gene encodes an NADPH-dependent beta-ketoacyl reductase which is specifically involved in rhamnolipid synthesis. J Bacteriol 180:4442–4451
5. Chen K, Wu HF, Grimsby J, Shih JC (1994) Cloning of a novel monoamine oxidase cDNA from trout liver. Mol Pharmacol 46:1226–1233
6. Cortes J, Haydock SF, Roberts GA, Bevitt DJ, Leadlay PF (1990) An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. Nature 348:176–178
7. Dhillon N, Hale RS, Cortes J, Leadlay PF (1989) Molecular characterization of a gene from *Saccharopolyspora erythraea* (*Streptomyces erythraeus*) which is involved in erythromycin biosynthesis. Mol Microbiol 3:1405–1414
8. Dhillon N, Leadlay PF (1990) A repeated decapeptide motif in the C-terminal domain of the ribosomal RNA methyltransferase from the erythromycin producer *Saccharopolyspora erythraea*. FEBS Lett 262:189–193
9. Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L (1991) Modular organization of genes required for complex polyketide biosynthesis. Science 252:675–679
10. Donadio S, Staver MJ (1993) IS1136, an insertion element in the erythromycin gene cluster of *Saccharopolyspora erythraea*. Gene 126:147–151
11. Finking R, Solsbacher J, Konz D, Schobert M, Schafer A, Jahn D, Marahiel MA (2002) Characterization of a new type of phosphopantetheinyl transferase for fatty acid and siderophore synthesis in *Pseudomonas aeruginosa*. J Biol Chem 277:50293–50302
12. Folcher M, Morris RP, Dale G, Salah-Bey-Hocini K, Viollier PH, Thompson CJ (2001) A transcriptional regulator of a pristinamycin resistance gene in *Streptomyces coelicolor*. J Biol Chem 276:1479–1485
13. French JC, Howells JD, Anderson LE (1970) Erythromycin process. US patent 3,551,294
14. Gaisser S, Bohm GA, Cortes J, Leadlay PF (1997) Analysis of seven genes from the *eryAI–eryK* region of the erythromycin biosynthetic gene cluster in *Saccharopolyspora erythraea*. Mol Gen Genet 256:239–251
15. Gaisser S, Bohm GA, Doumith M, Raynal MC, Dhillon N, Cortes J, Leadlay PF (1998) Analysis of *eryBI*, *eryBIII* and *eryBVII* from the erythromycin biosynthetic gene cluster in *Saccharopolyspora erythraea*. Mol Gen Genet 258:78–88
16. Green ED, Olson MV (1990) Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. Proc Natl Acad Sci USA 87:1213–1217
17. Grkovic S, Brown MH, Skurray RA (2002) Regulation of bacterial drug export systems. Microbiol Mol Biol Rev 66:671–701

18. Haydock SF, Dowson JA, Dhillon N, Roberts GA, Cortes J, Leadlay PF (1991) Cloning and sequence analysis of genes involved in erythromycin biosynthesis in *Saccharopolyspora erythraea*: sequence similarities between EryG and a family of *S*-adenosylmethionine-dependent methyltransferases. Mol Gen Genet 230:120–128
19. Hu Z, Pfeifer BA, Chao E, Murli S, Kealey J, Carney JR, Ashley G, Khosla C, Hutchinson CR (2003) A specific role of the *Saccharopolyspora erythraea* thioesterase II gene in the function of modular polyketide synthases. Microbiology 149:2213–2225
20. Jeevarajah D, Patterson JH, McConville MJ, Billman-Jacobe H (2002) Modification of glycopeptidolipids by an *O*-methyltransferase of *Mycobacterium smegmatis*. Microbiology 148:3079–3087
21. Kutchma AJ, Hoang TT, Schweizer HP (1999) Characterization of a *Pseudomonas aeruginosa* fatty acid biosynthetic gene cluster: purification of acyl carrier protein (ACP) and malonyl-coenzyme A:ACP transacylase (FabD). J Bacteriol 181:5498–5504
22. Miller ES (1991) Cloning vectors, mutagenesis, and gene disruption (*ermR*) for the erythromycin-producing bacterium *Aeromicrobium erythreum*. Appl Environ Microbiol 57:2758–2761
23. Miller ES, Woese CR, Brenner S (1991) Description of the erythromycin-producing bacterium *Arthrobacter* sp. strain NRRL B-3381 as *Aeromicrobium erythreum* gen. nov., sp. nov. Int J Syst Bacteriol 41:363–368
24. Mootz HD, Finking R, Marahiel MA (2001) 4′-phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis*. J Biol Chem 276:37289–37298
25. Nelson K, Paulsen I, Weinel C, Dodson R, Hilbert H, Fouts D, Gill S, Pop M, Martins Dos Santos V, Holmes M, Brinkac L, Beanan M, DeBoy R, Daugherty S, Kolonay J, Madupu R, Nelson W, White O, Peterson J, Khouri H, Hance I, Lee P, Holtzapple E, Scanlan D, Tran K, Moazzez A, Utterback T, Rizzo M, Lee K, Kosack D, Moestl D, Wedler H, Lauber J, Hoheisel J, Straetz M, Heim S, Kiewitz C, Eisen J, Timmis K, Duesterhoft A, Tummler B, Fraser C (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. Environ Microbiol 4:799–808
26. Paulus TJ, Tuan JS, Luebke VE, Maine GT, DeWitt JP, Katz L (1990) Mutation and cloning of *eryG*, the structural gene for erythromycin *O*-methyltransferase from *Saccharopolyspora erythraea*, and expression of *eryG* in *Escherichia coli*. J Bacteriol 172:2541–2546
27. Pereda A, Summers R, Katz L (1997) Nucleotide sequence of the *ermE* distal flank of the erythromycin biosynthesis cluster in *Saccharopolyspora erythraea*. Gene 193:65–71
28. Rawlings M, Cronan JE Jr (1992) The gene encoding *Escherichia coli* acyl carrier protein lies within a cluster of fatty acid biosynthetic genes. J Biol Chem 267:5751–5754
29. Reeves AR, Weber G, Cernota WH, Weber JM (2002) Analysis of an 8.1-kb DNA fragment contiguous with the erythromycin gene cluster of *Saccharopolyspora erythraea* in the *eryCI*-flanking region. Antimicrob Agents Chemother 46:3892–3899
30. Reeves AR, Cernota WH, Brikun IA, Wesley RK, Weber JM (2004) Engineering precursor flow for increased erythromycin production in *Aeromicrobium erythreum*. Metabol Eng DOI 10.1016/j.ymben.2004.03.003
31. Roberts AN, Hudson GS, Brenner S (1985) An erythromycin-resistance gene from an erythromycin-producing strain of *Arthrobacter* sp. Gene 35:259–270
32. Roberts AN, Barnett L, Brenner S (1987) Transformation of *Arthrobacter* and studies on the transcription of the *Arthrobacter ermA* gene in *Streptomyces lividans* and *Escherichia coli*. Biochem J 243:431–436
33. Salah-Bey K, Doumith M, Michel JM, Haydock S, Cortes J, Leadlay PF, Raynal MC (1998) Targeted gene inactivation for the elucidation of deoxysugar biosynthesis in the erythromycin producer *Saccharopolyspora erythraea*. Mol Gen Genet 257:542–553
34. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
35. Sanchez C, Du L, Edwards DJ, Toney MD, Shen B (2001) Cloning and characterization of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin. Chem Biol 8:725–738
36. Schwarzer D, Mootz HD, Linne U, Marahiel MA (2002) Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. Proc Natl Acad Sci USA 99:14083–14088
37. Stassi D, Donadio S, Staver MJ, Katz L (1993) Identification of a *Saccharopolyspora erythraea* gene required for the final hydroxylation step in erythromycin biosynthesis. J Bacteriol 175:182–189
38. Summers RG, Donadio S, Staver MJ, Wendt-Pienkowski E, Hutchinson CR, Katz L (1997) Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of *Saccharopolyspora erythraea* that are involved in L-mycarose and D-desosamine production. Microbiology 143:3251–3262
39. Thompson CJ, Kieser T, Ward JM, Hopwood DA (1982) Physical analysis of antibiotic-resistance genes from *Streptomyces* and their use in vector construction. Gene 20:51–62
40. Uchiyama H, Weisblum B (1985) *N*-Methyl transferase of *Streptomyces erythraeus* that confers resistance to the macrolide-lincosamide-streptogramin B antibiotics: amino acid sequence and its homology to cognate R-factor enzymes from pathogenic bacilli and cocci. Gene 38:103–110
41. Valdez F, Gonzalez-Ceron G, Kieser HM, Servin-Gonzalez L (1999) The *Streptomyces coelicolor A3(2) lipAR* operon encodes an extracellular lipase and a new type of transcriptional regulator. Microbiology 145:2365–2374
42. van den Boom TJ (2000) Recent developments in the molecular genetics of the erythromycin-producing organism *Saccharopolyspora erythraea*. Adv Appl Microbiol 47:79–111
43. Vara J, Lewandowska-Skarbek M, Wang YG, Donadio S, Hutchinson CR (1989) Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea (Streptomyces erythreus)*. J Bacteriol 171:5872–5881
44. Weber JM, Wierman CK, Hutchinson CR (1985) Genetic analysis of erythromycin production in *Streptomyces erythreus*. J Bacteriol 164:425–433
45. Weber JM, Schoner B, Losick R (1989) Identification of a gene required for the terminal step in erythromycin A biosynthesis in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). Gene 75:235–241
46. Weber JM, Leung JO, Maine GT, Potenz RH, Paulus TJ, DeWitt JP (1990) Organization of a cluster of erythromycin genes in *Saccharopolyspora erythraea*. J Bacteriol 172:2372–2383
47. Weber JM, Leung JO, Swanson SJ, Idler KB, McAlpine JB (1991) An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea*. Science 252:114–117
48. Weissman KJ, Hong H, Oliynyk M, Siskos AP, Leadlay PF (2004) Identification of a phosphopantetheinyl transferase for erythromycin biosynthesis in *Saccharopolyspora erythraea*. Chem Biochem 5:116–125
49. Xue Y, Zhao L, Liu HW, Sherman DH (1998) A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity. Proc Natl Acad Sci USA 95:12111–12116