Supporting information

Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites

MAMORU KOMATSU[1], KYOKO KOMATSU[1], HANAE KOIWAI[1], YUUKI YAMADA[1], IKUKO KOZONE[2], MIHO IZUMIKAWA[2], JUNKO HASHIMOTO[2], MOTOKI TAKAGI[2], SATOSHI OMURA[3], KAZUO SHIN-YA[4], DAVID E. CANE[5] and HARUO IKEDA[1]

[1]Kitasato Institute for Life Sciences, Kitasato University, 1-15-1 Kitasato, Minami-ku, Sagamihara, Kanagawa 252-0373, Japan, [2]Japan Biological Informatics Consortium, 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan, [3]Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan, [4]National Institute of Advanced Industrial Science, and Technology, 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan and [5]Department of Chemistry, Box H, Brown University, Providence, RI 02912-9108, USA.

SI METHODS

Construction of genomic library. To prepare the cosmid library, actinomycete microorganisms were grown in Tryptic Soy Broth (TSB) at 30°C for 2 to 3 days with agitation. Mycelia were digested with egg-white lysozyme (2 mg/mL) at 37°C for 5 to 20 min, and then resulting protoplasts were lysed by addition of sodium lauryl sulfate (1%). The lysate was clarified by treatment with proteinase K (0.5 mg/mL) at 45°C for 1 hr. After removal of undigested proteins and lipids by repeated phenol-chloroform extraction, chromosomal DNA was purified in 2 rounds of

precipitation steps using 2-propanol and polyethylene glycol 6,000, respectively. The final precipitate was washed with 70% ethanol and washed DNA was dissolved in TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) to make about 500 μg/mL. The chromosomal DNA preparation was sheared by repeated passing through a 23-gauge needle (300 to 500 times) to generate about 40-kb fragments. After DNA fragments had been separated by field-inversion gel electrophoresis (the time setting was 1.2 sec for forward and 0.4 sec for reverse directions) in 0.8% low-melting agarose gel, both edges of the gel were cut and stained in ethidium bromide. The DNA size was confirmed, marked and the corresponding area of the remaining non-stained gel containing 40-kb fragments was cut and melted by heating at 65°C. The melted solution was extracted with an equal volume of 50 mM Tris-HCl (pH 8.0) -saturated phenol to remove agarose as precipitates between aqueous and organic phases. The aqueous upper phase containing DNA fragments was concentrated by extraction with 1-butanol and then the concentrate was dialyzed against TE. Purified DNA fragments were treated with Bal31 for 15 sec at 25°C and both ends of fragments were filled with T4 DNA polymerase and T4 polynucleotide kinase in the presence of dNTPs at 37°C for 15 min. Blunt-ended DNA fragments were ligated with the large *Swa*I segment of pKU402[1] and the ligated DNA was subjected to *in vitro* packaging. Transfectants were selected by resistance to 100 μg/mL ticarcillin. Clones carrying the entire biosynthetic gene cluster for the secondary metabolite were confirmed by the detection of two regions, upstream and downstream of the gene cluster by PCR amplification. The cosmid clone selected was digested with *Ase*I to generate about 40-kb insert and the insert was subcloned into the *Nde*I site of pKU465cos[2] by *in vitro* packaging. Desired subclones were selected by resistance to 50 μg/mL kanamycin.

To prepare a BAC library, mycelia grown in TSB containing 5 mM $MgCl_2$ were embedded in 0.5 % InCert® agarose (Lonza, ME, USA) before digestion with 1 mg/mL lysozyme (30°C for at

least 6 hr). The resulting protoplasts generated were lysed by addition of 0.5% sodium N-lauroylsarcosinate and 1 mg/mL proteinase K in 50 mM Tris-HCl (pH 9.5) at 50°C for 24 h. All supernatant was removed and proteinase K was inactivated by 0.1 mM (*p*-amidinophenyl) methanesulfonyl fluoride hydrochloride or 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride. After removal of the detergent by repeated washing with 50 mM EDTA (pH 8.0), DNA embedded in the agarose plug was partially digested with *Bam*HI. About 100 to 150-kb fragments were purified by CHEF electrophoresis. Agarose gel containing DNA fragments was melted by NaI at half the gel weight. Melted agarose and NaI were removed by dialysis (0.025 μm pore size filter) against TE. The residual agarose was completely digested by DNase-free agarase. Purified DNA fragments were ligated with a large *Bam*HI segment of pKU503[2] and *E. coli* DH10B or NEB 10-beta was transformed by electroporation with ligated DNA. The desired clone was selected by the insert size and amplification of two regions, upstream and downstream of the gene cluster, by PCR amplification.

***In vivo* subcloning of the entire biosynthetic gene cluster**. About 250 to 500-bp homologous regions upstream and downstream of the gene cluster were prepared by PCR amplification with the cosmid clone carrying the entire biosynthetic gene cluster for the secondary metabolite as a template DNA using two primer pairs corresponding to upstream and downstream of the gene cluster. Linearized cloning vector was prepared by PCR amplification with pRED[2] lacking the *Xba*I site as a template DNA using the primer pair, forward: 5'-CTCGAG*TCTAGA*<u>TGCCAGGAAGATACTTAACAG</u>-3' (italicized and underlined letters indicate the *Xba*I site and the region of pRED, respectively) or 5'-CTCGAG*ACTAGT*<u>TGCCAGGAAGATACTTAACAG</u>-3' (italicized and underlined letters indicate the *Spe*I site and the region of pRED, respectively) and reverse:

5’-CTCGAG*AAGCTT*CCATTCATCCGCTTATTATC-3’ (italicized and underlined letters indicate the *Hin*dIII site and the region of pRED, respectively). Initial denaturation at 96°C for 180 sec was followed by 25 cycles of amplification (95°C 30 sec, 60 °C 30 sec and 72°C 30 sec) and final incubation at 72°C for 5 min using the Expand High Fidelity PCR system (Roche Diagnostics, Tokyo, Japan) or Phusion DNA polymerase (New England Biolabs, MA, USA). After removal of template DNA by treatment with *Dpn*I, each amplicon was digested with *Xba*I/*Hin*dIII, *Spe*I/*Hin*dIII or *Bam*HI/*Hin*dIII. Both upstream and downstream regions of the biosynthetic gene cluster for pholipomycin of *S. clavuligerus* ATCC 27064 were prepared by PCR with the cosmid clone carrying the entire biosynthetic gene cluster for pholipomycin using two primer pairs, upstream forward: 5’-CTCGAG*TCTAGA*CTCGTGATGGGAACGGAGTTCGAC-3’ (italicized and underlined letters indicate the *Xba*I site and C-terminal region of *sclav_p1272*, respectively) and upstream reverse: 5’-CCCTGAT*TGTAC*AGGAGCTTCAGCGCTTCCTGGAG-3’ (italicized and underlined letters indicate the *Bsr*GI site and N-terminal region of *sclav_p1273*, respectively), and downstream forward: 5’-AAGCTCC*TGTACA*ATCAGGGAAGAGGGGAAGGGGAAG-3’ (italicized and underlined letters indicate the *Bsr*GI site and the region between *sclav_p1291* and *sclav_p1292*, respectively) and downstream reverse: 5’-CTCGAG*AAGCTT*GGAAACATGAGCTGCACCGAACC-3’ (italicized and underlined characters indicate the *Hin*dIII site and C-terminal region of *sclav_p1292*, respectively). Both upstream and downstream regions of the biosynthetic gene cluster for holomycin of *S. clavuligerus* ATCC 27064 were prepared by PCR with the cosmid clone carrying the gene cluster for holomycin biosynthesis using two primer pairs, upstream forward: 5’-CTCGAG*ACTAGT*CGATCTTGTCCGCTACCGGCTCTA-3’ (italicized and underlined letters indicate *Spe*I site and N-terminal region of *sclav_5265*, respectively) and upstream reverse:

5’-GGAAGTAGA*TGTACA*TCCGGAACTTGTAGCACCCCGTGAAC-3’ (italicized and underlined letters indicate the *Bsr*GI site and region in *sclav_5266*, respectively), and downstream forward: 5’-CGGA*TGTACA*TCACCGTGCGATCCACTCCTCCAG-3’ (italicized and underlined letters indicate the *Bsr*GI site and region in *sclav_5281*, respectively) and downstream reverse: 5’-CTAGAG*AAGCTT*AGTCCTTTCGCGGTCTTCCCTGTC-3’ (italicized and underlined letters indicate the *Hin*dIII site and the region between *sclav_5281* and *sclav_5282*, respectively), respectively. Both upstream and downstream regions of the biosynthetic gene cluster for clavulanic acid of *S. clavuligerus* ATCC 27064 were prepared by PCR with the cosmid clone carrying gene cluster for clavulanic acid biosynthesis using two primer pairs, upstream forward: 5’-CTCGAG*GGATCC*TCATCGCCATAGTTGTCGGAATGC-3’ (italicized and underlined letters indicate the *Bam*HI site and N-terminal region of *sclav_4198*, respectively) and upstream reverse: 5’-CTCGAG*ATGCAT*GTTTTGTCGCTACTCCCCGCTGTC-3’ (italicized and underlined letters indicate the *Nsi*I site and the region between *sclav_4197* and *sclav_4198*, respectively), and downstream forward: 5’-CTCGAG*ATGCAT*GGACCCCCAGGACAGGATACGAAG-3’ (italicized and underlined letters indicate the *Nsi*I site and the region between *sclav_4179* and *sclav_4180*, respectively) and downstream reverse: 5’-CTCGAG*AAGCTT*GTACGGAAGTGACGGAGCCCACAG-3’ (italicized and underlined letters indicate the *Hin*dIII site and the region between *sclav_4179* and *sclav_4180*, respectively). Both upstream and downstream regions of the biosynthetic gene cluster for chloramphenicol of *S. venezuelae* ATCC 10712 were prepared by PCR with cosmid clone pAH91 using two primer pairs, upstream forward: 5’-CTCGAG*ACTAGT*CCACGTACACCCTGCTCCTCACC-3’ (italicized and underlined letters indicate the *Spe*I site and the region in *sven_0933*, respectively) and upstream reverse: 5’-CGCGA*TGTACA*TCCTCCTGGAGTCACTTGCCGCATC-3’ (italicized and underlined

letters indicate the *Bsr*GI site and the region between *sven_0932* and *sven_0933*, respectively), and downstream forward: 5'-GCGTGGA*TGTACA*TCGCACCATGCGTTCGCGATCCTTTCG-3' (italicized and underlined letters indicate the *Bsr*GI site and C-terminal region of *sven_0912*, respectively) and downstream reverse: 5'-GTCGAG*AAGCTT*CACCCCGAAGAACTGGGCGAGGAC-3' (italicized and underlined letters indicate the *Hin*dIII site and the region in *sven_0911*, respectively). The initial denaturation step (at 96°C for 180 sec) was followed by 25 cycles of amplification (95°C 30 sec, 60 °C 30 sec and 72°C 90 sec) and then final incubation at 72°C for 5 min. All amplicons were treated with *Dpn*I to remove the template DNA. Amplicons of upstream regions were then digested with *Xba*I (pholipomycin cluster), *Spe*I (holomycin and chloramphenicol clusters) or *Bam*HI (clavulanic acid cluster), and *Bsr*GI (pholipomycin, holomycin and chloramphenicol clusters) or *Nsi*I (clavulanic acid cluster). All amplicons of downstream regions were digested with *Bsr*GI or *Nsi*I (amplicon of clavulanic acid cluster), and *Hin*dIII. The amplicon of pRED lacking the *Xba*I site digested with *Xba*I/*Hin*dIII (for pholipomycin cluster), *Spe*I/*Hin*dIII (for holomycin and chloramphenicol clusters) or *Bam*HI/*Hin*dIII (for clavulanic acid cluster), and each upstream and downstream amplicon were ligated together in a molar ratio of 1:1:1. *E. coli* DH5α was transformed with the ligated products, the transformants were selected by resistance to 30 μg/mL chloramphenicol and the desired clones were confirmed by restriction digestion and PCR amplification. The resulting recombinant plasmids carrying upstream and downstream regions of the gene cluster were linearized by PCR amplification using the primer pair, and downstream forward and upstream reverse primers. Initial denaturation (96°C for 180 sec) was followed by 25 cycles of amplification (95°C 30 sec, 60 °C 30 sec and 72°C 160 sec) and final incubation at 72°C for 5 min. The amplicon was treated with *Dpn*I to remove the circular template DNA and digested with Bal31 for 30 sec at

30°C. Alternatively, when the biosynthetic gene cluster was less than 15 kb, the entire biosynthetic gene cluster was subcloned by homologous recombination without preparing upstream and downstream amplicons of the gene cluster. Since the gene cluster for lactacystin biosynthesis was about 13 kb, pRED was directly linearized by PCR with pRED vector using the designated primer pair containing both 41 nt sequences corresponding to upstream and downstream regions of the biosynthetic gene cluster for lactacystin, forward: 5'-TCGAGGTGACGCGGGTGGTCGCGATTCAGCGGCTCACTGTG*GAATTC***TGCCAGGAA GATACTTAACAG**-3' (underlined, italicized and bold letters indicate upstream region of the biosynthetic gene cluster for lactacystin of *S. lactacystinaeus* OM-6519, the *Eco*RI site and the region of vector pRED, respectively) and reverse: 5'-TGGGACGCACCGTCTACCTCGAACCGCAGGAGCAGGAGGCG*GAATTC***CCATTCATCC GCTTATTATC**-3' (underlined, italicized and bold letter indicate downstream region of the biosynthetic gene cluster for lactacystin of *S. lactacystinaeus* OM-6519, the *Eco*RI site and the region of vector pRED, respectively). Initial denaturation (96°C for 180 sec) was followed by 5 cycles of amplification (95°C 30 sec, 50 °C 30 sec and 72°C 100 sec), 25 cycles (95°C 30 sec and 68°C 100 sec) and then final incubation at 72°C for 5 min. After amplification, the reaction mixture was treated with *Dpn*I to remove template DNA.

The biosynthetic gene cluster for the secondary metabolite in the cosmid clone was replaced with the linearized vector carrying upstream and downstream regions of the gene cluster by homologous regions between both plasmids using the λ-RED recombination system[3]. The amplicon of pRED carrying upstream and downstream regions of the gene cluster was co-transformed L-arabinose-induced *E. coli* DH10B carrying pKD46[3] or pKD119 (http://cgsc.biology.yale.edu/) with the linearized (*Ase*I-digestion) cosmid clone carrying the entire biosynthetic gene cluster by

electroporation, as described previously[4]. The desired plasmids, pRED carrying the entire biosynthetic gene cluster, were obtained by selection with 30 μg/mL chloramphenicol and were also confirmed by restriction digestion. After purification of the large segment of *Xba*I/*Hin*dIII (pholipomycin cluster), *Spe*I/*Hin*dIII (holomycin and chloramphenicol clusters), *Bam*HI/*Nsi*I (clavulanic acid cluster) or *Eco*RI (lactacystin cluster) of the pRED clone carrying the entire biosynthetic gene cluster by agarose gel electrophoresis, the fragments containing the entire gene cluster were ligated with *Xba*I/*Hin*dIII-pKU492A*aac(3)IV* (the vector was derived from pKU460[2] by replacement of two regions, ϕC31 *attP-int* to TG1 *attP-int*[5] and *aphII* to *loxP-aac(3)IV-loxP*) (pholipomycin, holomycin and chloramphenicol clusters), *Bam*HI/*Pst*I-pKU492A*aac(3)IV* (clavulanic acid cluster) or *Eco*RI (lactacystin cluster) -pKU492 to generate pPHM1 (pholipomycin cluster), pHLM1 (holomycin cluster), pCML1 (chloramphenicol cluster), pCLV1 (clavulanic acid cluster) or pLTC1 (lactacystin cluster), respectively. To express the biosynthetic gene cluster for clavulanic acid by introducing a gene encoding the transcriptional regulator, *ccaR* with its own promoter was cloned by PCR with a cosmid clone containing the biosynthetic gene cluster for cephamycin C[2] using the primer pair, forward: 5'-CG*GGATCC*TAGGGAGGGGAGAGTCCGACAGCC-3' (italicized and underlined letters indicate the *Bam*HI site and promoter region of *ccaR*, respectively) and reverse: 5'-CTCGAG*GCTAGC*GGTGGTTTTCGGTCTGCTCTGTGTGAGG-3' (italicized and underlined letters indicate the *Nhe*I site and the region between *sclav_4203* and *ccaR*, respectively). The PCR program employed was as follows: initial denaturation step (95°C, 180 sec) was followed by 25 cycles of amplification (95°C 30 sec, 60°C 30 sec and 72°C 100 sec) and then final incubation at 72°C for 5 min. The reaction mixture was treated with *Dpn*I to remove template DNA and the amplicon was digested with *Bam*HI and *Nhe*I. The resultant 1.3-kb *Bam*HI/*Nhe*I fragment

containing *ccaR* was ligated with the large fragment of *Bam*HI/*Nhe*I-pCLV1 to generate pCLV1*. Since no regulatory genes are located in the biosynthetic gene cluster for lactacystin, the gene cluster was controlled by an alternative promoter. The promoter region of *rpsJ* gene encoding ribosomal protein S10 was prepared by PCR with pKU460::*rpsJp*[6] using the primer pair, forward: 5'-*GTGACGGCGGATATGTCTCGCGCTGCGCGCTCTTGGGG**CAT***ATGTACTCAGTAGTCCTT CGTCTCG-3' (italicized and underlined letters indicate the N-terminal region of *ltcA* gene encoding lactacystin biosynthesis and the region of *rpsJ* promoter with ribosome binding site, respectively, and bold letters correspond to the start codon of *ltcA*) and reverse: 5'-*AGCAGCCCTTGCGCCCTGAGTGCTTGCGGCAGCGTGAAGCTAGC*GATCTCGGCTTGAAC GAATTG-3' (italicized and underlined letters indicate the region of vector pKU492 and downstream region of *aac(3)I* of pKU460::*rpsJp*, respectively). Initial denaturation (96°C for 180 sec) was followed by 5 cycles of amplification (95°C 30 sec, 50°C 30 sec and 72°C 90 sec), 25 cycles (95°C 30 sec and 68°C 100 sec) and then final incubation at 72°C for 5 min. The reaction mixture was then treated with *Dpn*I to remove the template DNA. The amplicon containing *rpsJp* and a resistance marker, *aac(3)I*, was co-transformed with L-arabinose-induced *E. coli* BW25113[3] carrying pKD119 with linearized pLTC1 (*Bsr*GI-digestion). The desired plasmid, pLTC1* (expression of the biosynthetic gene cluster for lactacystin was controlled by *rpsJ* promoter), was obtained by selection with 50 μg/mL kanamycin and 25 μg/mL fortimicin (*aac(3)I*), and was confirmed by restriction digestion.

Desired plasmids carrying the minimum size of the entire biosynthetic gene cluster were transferred to *E. coli* GM2929 *hsdS*::Tn*10* to prepare unmethylated DNA preparation and the unmethylated recombinant plasmids were introduced into *S. avermitilis* SUKA17 or SUKA22 by polyethylene glycol-assisted protoplast transformation[2].

Construction of cassettes for expression of plant diterpene synthase genes. A synthetic *tds* and *lps* genes encoding the taxa-4,11-diene synthase of *Taxus brevifolia* (Q41594) and levopimaradiene synthase of *Ginkgo biloba* (Q947C4) were synthesized (OPERON Biotechnologies K.K., Tokyo Japan) according to *Streptomyces* codon usage and a 27-bp sequence (5'-**TCTAGA**CTCGACA*GGAGGA*GCCATATG-3'; bold, italicized and underlined letters indicate the *Xba*I site, start codon of the synthase gene and ribosome binding site, respectively) and the *Hin*dIII site were added at the 5′ and 3′ ends, respectively. Each synthetic gene was introduced into the large *Xba*I-*Hin*dIII fragment of pKU1021::*crtE*[7]. After confirmation of each recombinant plasmid by restriction digestion, unmethylated DNA preparation in *E. coli* GM2929 *hsdS*::Tn*10* was used for the transformation of *S. avermitilis*.

SI REFERENCES

1. Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M. (2001) Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites *Proc. Natl. Acad. Sci. USA* ***98***, 12215-12220.
2. Komatsu, M., Uchiyama, T., Omura, S., Cane, D.E. and Ikeda, H. (2010) Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. USA*. ***107***, 2646-2651.
3. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. USA.* ***97***, 6640-6645.
4. Jiang, J., Tetzlaff, C.N., Takamatsu, S., Iwatsuki, M., Komatsu, M., Ikeda, H. and Cane, D.E. (2009) Genome mining in *Streptomyces avermitilis*: A biochemical Baeyer-Villiger reaction and

discovery of a new branch of the pentalenolactone family tree. *Biochemistry 48*, 6431-6440.

5. Komatsu, M., Tsuda, M., Omura, S., Oikawa, H. and Ikeda, H. (2008) Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol. *Proc. Natl. Acad. Sci. USA*. *105*, 7422-7427.
6. Morita, K., Yamamoto, T., Fusada, N., Komatsu, M., Ikeda, H., Hirano, N. and Takahashi, H. (2009) The site-specific recombination system of actinophage TG1. *FEMS Microbiol. Lett.*, ***297***, 234-240.
7. Yamada, Y., Cane, D.E. and Ikeda, H. (2012) Diversity and Analysis of Bacterial Terpene Synthases. *In Chapter seven, Natural Product Biosynthesis by Microorganisms and Plants, Part A, Methods in Enzymology, Volume 515* (D.A. Hopwood ed.), pp. 123-166, Elsevier Inc. Academic Press, New York.

SI FIGURE LEGEND

Figure S1. Maps of biosynthetic gene clusters for (A) streptomycin, (B) ribostamycin, (C) kasugamycin, (D) pholipomycin, (E) oxytetracycline, (F) resistomycin, (G) cephamycin C, (H) holomycin, (I) lactacystin, (J) clavulanic acid, (K) rebeccamycin, (L) novobiocin, (M) chloramphenicol, (N) 2-methylisoborneol, (O) pladienolide, (P) erythromycin, (Q) bafilomycin, (R) nemadectin, (S) aureothin, and (T) leptomycin. Maps from (O) to (T) are drawn to a different scale.