

# 広島大学学術情報リポジトリ

## Hiroshima University Institutional Repository

Title	The genome sequence of <i>Streptomyces rochei</i> 7434AN4, which carries a linear chromosome and three characteristic linear plasmids
Auther(s)	Nindita, Yosi; Cao, Zhisheng; Fauzi, Amirudin Akhmad; Teshima, Aiko; Misaki, Yuya; Muslimin, Rukman; Yang, Yingjie; Shiwa, Yuh; Yoshikawa, Hirofumi; Tagami, Michihira; Lezhava, Alexander; Ishikawa, Jun; Kuroda, Makoto; Sekizuka, Tsuyoshi; Inada, Kuninobu; Kinashi, Haruyasu; Arakawa, Kenji
Citation	Scientific Reports , 9 : 10973
Issue Date	2019-07-29
DOI	<a href="https://doi.org/10.1038/s41598-019-47406-y">10.1038/s41598-019-47406-y</a>
Self DOI	
URL	<a href="http://ir.lib.hiroshima-u.ac.jp/00048194">http://ir.lib.hiroshima-u.ac.jp/00048194</a>
Right	© The Author(s) 2019. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a

	copy of this license, visit <a href="http://creativecommons.org/licenses/by/4.0/">http://creativecommons.org/licenses/by/4.0/</a> .
Relation	



OPEN

# The genome sequence of *Streptomyces rochei* 7434AN4, which carries a linear chromosome and three characteristic linear plasmids

Yosi Nindita<sup>1,2</sup>, Zhisheng Cao<sup>1</sup>, Amirudin Akhmad Fauzi<sup>1</sup>, Aiko Teshima<sup>1</sup>, Yuya Misaki<sup>1,2</sup>, Rukman Muslimin<sup>1</sup>, Yingjie Yang<sup>1</sup>, Yuh Shiwa<sup>3</sup>, Hirofumi Yoshikawa<sup>3,4</sup>, Michihira Tagami<sup>5</sup>, Alexander Lezhava<sup>5</sup>, Jun Ishikawa<sup>6</sup>, Makoto Kuroda<sup>7</sup>, Tsuyoshi Sekizuka<sup>7</sup>, Kuninobu Inada<sup>8</sup>, Haruyasu Kinashi<sup>1</sup> & Kenji Arakawa<sup>1,2</sup>

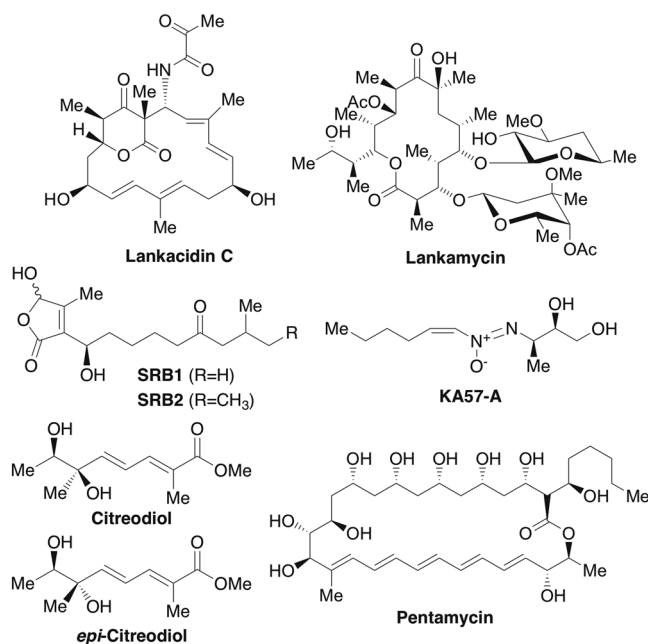
*Streptomyces rochei* 7434AN4 produces two structurally unrelated polyketide antibiotics, lankacidin and lankamycin, and carries three linear plasmids, pSLA2-L (211 kb), -M (113 kb), and -S (18 kb), whose nucleotide sequences were previously reported. The complete nucleotide sequence of the *S. rochei* chromosome has now been determined using the long-read PacBio RS-II sequencing together with short-read Illumina Genome Analyzer IIx sequencing and Roche 454 pyrosequencing techniques. The assembled sequence revealed an 8,364,802-bp linear chromosome with a high G + C content of 71.7% and 7,568 protein-coding ORFs. Thus, the gross genome size of *S. rochei* 7434AN4 was confirmed to be 8,706,406 bp including the three linear plasmids. Consistent with our previous study, a *tap-tpg* gene pair, which is essential for the maintenance of a linear topology of *Streptomyces* genomes, was not found on the chromosome. Remarkably, the *S. rochei* chromosome contains seven ribosomal RNA (*rnn*) operons (16S-23S-5S), although *Streptomyces* species generally contain six *rnn* operons. Based on 2ndFind and antiSMASH platforms, the *S. rochei* chromosome harbors at least 35 secondary metabolite biosynthetic gene clusters, including those for the 28-membered polyene macrolide pentamycin and the azoxyalkene compound KA57-A.

The filamentous Gram-positive soil bacterial genus *Streptomyces* is well characterized by its prolific potential to produce a vast array of secondary metabolites, including agriculturally and clinically useful antibiotics. Unlike other bacteria, their chromosomes are linear with a size of around 8–9 Mb. *Streptomyces* linear replicons harbor terminal inverted repeat (TIR) sequences at both ends and the 5'-ends are covalently bound to terminal protein (TP)<sup>1</sup>. *Streptomyces* linear chromosomes frequently undergo spontaneous deletions, leading to DNA rearrangements including amplification, arm replacement, and circularization<sup>2–4</sup>.

<sup>1</sup>Department of Molecular Biotechnology, Graduate School of Advanced Sciences of Matter, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, 739-8530, Japan. <sup>2</sup>Unit of Biotechnology, Division of Biological and Life Sciences, Graduate School of Integrated Sciences for Life, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, 739-8530, Japan. <sup>3</sup>NODAI Genome Research Center, Tokyo University of Agriculture, 1-1-1 Sakuragaoka, Setagaya-ku, Tokyo, 156-8502, Japan. <sup>4</sup>Department of Bioscience, Tokyo University of Agriculture, 1-1-1 Sakuragaoka, Setagaya-ku, Tokyo, 156-8502, Japan. <sup>5</sup>Omics Science Center, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan. <sup>6</sup>Department of Bioactive Molecules, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo, 162-8640, Japan. <sup>7</sup>Pathogen Genomics Center, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo, 162-8640, Japan. <sup>8</sup>Natural Science Center for Basic Research and Development, Hiroshima University, 1-4-2 Kagamiyama, Higashi-Hiroshima, 739-8526, Japan. Correspondence and requests for materials should be addressed to K.A. (email: [karakawa@hiroshima-u.ac.jp](mailto:karakawa@hiroshima-u.ac.jp))

Species <sup>a</sup>	Length (bp)	G + C content (%)	CDS (no.)	Average CDS length (bp)	TIR (bp)	rRNAs (no.)	tRNA (no.)
<i>S. rochei</i> 7434AN4	8,364,802	71.7	7,568	973	53,892	7	67
<i>S. coelicolor</i> A3(2)	8,667,507	72.1	7,825	991	21,653	6	63
<i>S. avermitilis</i> MA-4680	9,025,608	70.7	7,582	1,027	49	6	68
<i>S. griseus</i> IFO13350	8,545,929	72.2	7,138	1,055	132,910	6	66
<i>S. hygroscopicus</i> 5008	10,145,833	71.9	8,849	952	14	6	68
<i>S. scabies</i> 87.22	10,148,695	71.5	8,746	1,005	18,488	6	75

**Table 1.** General features of the chromosomes of *Streptomyces rochei* 7434AN4 and five other *Streptomyces* species. <sup>a</sup>GenBank accession number: *S. rochei* 7434AN4 chromosome, AP018517; *S. coelicolor* A3(2), NC\_003888; *S. avermitilis* MA-4680, NC\_003155; *S. griseus* IFO13350, NC\_010572; *S. hygroscopicus* 5008, CP003275; *S. scabies* 87.22, NC\_013929.



**Figure 1.** Secondary metabolites produced by *Streptomyces rochei* 7434AN4 and its mutants. Polyketide antibiotics; lankacidin C and lankamycin. Signaling molecules; SRB1 and SRB2. Azoxyalkene; KA57-A. Polyketides; citreodiol, *epi*-citreodiol, and pentamycin.

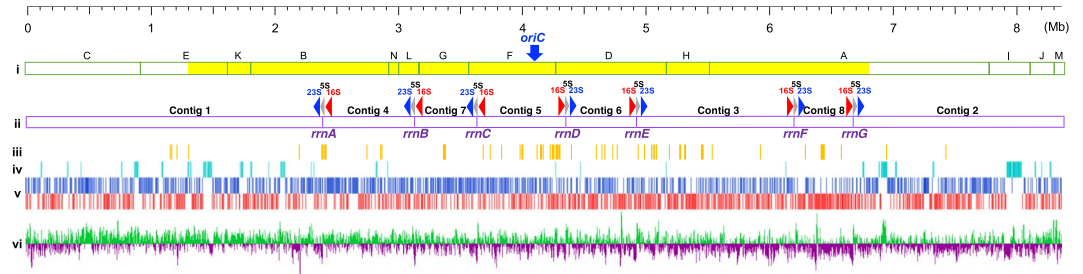
To date, over 1,141 *Streptomyces* strains have been sequenced and deposited in the GenBank database (<ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/>) (as of 26th of June 2019), including the actinorhodin producer *Streptomyces coelicolor* A3(2)<sup>5</sup>, the avermectin producer *Streptomyces avermitilis*<sup>6</sup>, and the streptomycin producer *Streptomyces griseus*<sup>7</sup> (Table 1). *Streptomyces* species have a great potential to produce over 20 secondary metabolites, including polyketides, non-ribosomal peptides, terpenoids, aminoglycosides, siderophores, and others. However, most of these biosynthetic gene clusters are poorly expressed or are not at all under normal culture conditions. Thus, *Streptomyces* genomes are a valuable source for natural product discovery.

*Streptomyces rochei* 7434AN4 produces two structurally unrelated polyketides, lankacidin C and lankamycin (Fig. 1)<sup>8,9</sup>, and carries three characteristic linear plasmids (pSLA2-L, -M, and -S), whose nucleotide sequences were previously determined (Table 2). The 210,614-bp largest linear plasmid, pSLA2-L (Accession number; AB088224), carries 143 open reading frames (ORFs), including the biosynthetic gene clusters for lankacidin, lankamycin, uncharacterized type-II polyketide, and carotenoid<sup>10</sup>. This plasmid carries many regulatory genes and the biosynthetic gene for the signaling molecules SRBs (*Streptomyces rochei* butenolides) (Fig. 1) that induce lankacidin and lankamycin production in *S. rochei*<sup>11</sup>. The 113,464-bp linear plasmid pSLA2-M (AB597522) comprises of 121 ORFs and carries several self-defense genes including a CRISPR (clustered regularly interspaced short palindromic repeats) cluster and a *ku70/ku80*-like gene<sup>12</sup>. Both plasmids harbor a *tap-tpg* gene pair that encodes a telomere-associated protein and a TP necessary for end patching of linear replicons<sup>13</sup>. The 17,526-bp smallest plasmid pSLA2-S (AB905437) consists of 17 ORFs and does not contain a *tap-tpg* gene pair in contrast to pSLA2-L and -M. Basic features of the three linear plasmids are shown in Table 2.

Regarding the genomic structures in *S. rochei* 7434AN4, the right end sequences of pSLA2-L and -M are almost identical (99.9%) up to 14.6 kb from the end<sup>12</sup>. Partial sequencing and Southern blot analysis revealed

Replicons <sup>a</sup>	Length (bp)	G + C content (%)	CDS (no.)	Average CDS length (bp)	TIR (bp)
Chromosome	8,364,802	71.7	7,568	973	53,892
pSLA2-L	210,614	72.8	143	1,330	1,992
pSLA2-M	113,464	69.7	121	718	352
pSLA2-S	17,526	69.7	22	508	817

**Table 2.** General features of the chromosome and three plasmids of *S. rochei* 7434AN4. <sup>a</sup>GenBank accession number: *S. rochei* 7434AN4 chromosome, AP018517; pSLA2-L, AB088224; pSLA2-M, AB597522; pSLA2-S, AB905437.



**Figure 2.** Schematic representation of the *S. rochei* chromosome. Scale bars are drawn in megabases. (i) *AseI* physical map. The possible core region of the *S. rochei* chromosome (1.31–6.80 Mb) is marked as yellow. (ii) Distribution of rRNA-encoding gene (rDNA) operons and eight assembled contigs. Gray and blue arrows are 5S-23S rDNA operon, while red is 16S rDNA. (iii) Distribution of tRNAs. (iv) Distribution of secondary metabolite gene clusters. (v) Distribution of CDSs according to direction of transcription (+ strand, upper line; – strand, lower line). (vi) GC-skew for 10-kb window and 500-bp step. The putative *oriC* (gene) locus is indicated by a blue arrow. Main features were generated by DNA plotter software (<https://www.sanger.ac.uk/science/tools/dnaplotter>)<sup>15</sup>.

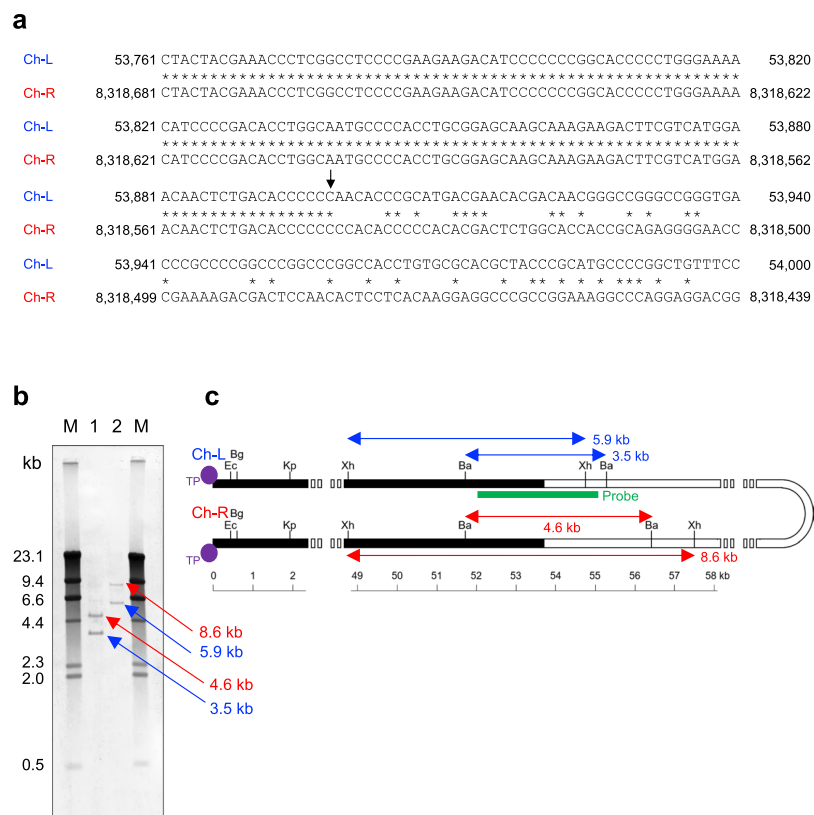
that both ends of the chromosome are identical to each other, and share 98.5% homology with the right end of pSLA2-L and -M up to 3.1 kb. Furthermore, a truncated *tpg* homolog was detected in one contig<sup>14</sup>. In addition, curing of pSLA2-L from strain 51252, which harbors only pSLA2-L, caused terminal deletions of the chromosome followed by circularization in mutant 2–39<sup>9,14</sup>. These results suggest that the *tap-tpg* of pSLA2-L or -M functions for terminal replication to maintain a linear topology of the chromosome. This hypothesis was supported by complementation and curing experiments of the *tap-tpg* of pSLA2-M<sup>14</sup>. However, the absence of a *tap-tpg* pair on the chromosome still remained to be proved by genome sequencing.

*Streptomyces* species are well known for their three characteristics: possession of linear replicons, complex morphological differentiation, and an ability to produce secondary metabolites. In this study, we have determined the complete nucleotide sequence of the linear chromosome of *Streptomyces rochei* 7434AN4 and extensively analyzed these three characteristics in comparison with other *Streptomyces* strains hitherto characterized.

## Results and Discussion

### Nucleotide sequencing and physical analysis of the linear chromosome of *S. rochei* 7434AN4.

The complete nucleotide sequence of the linear chromosome of *S. rochei* 7434AN4 was obtained by assembling a combination of reads from the long-read PacBio RS-II sequencing together with short-read Illumina Genome Analyzer IIx (GAIIx) sequencing and Roche 454 pyrosequencing. The 26,119,215 trimmed reads (217-fold coverage of the whole genome) obtained through Illumina GAIIx sequencing were assembled using ABySS protocol to give 340 contigs with >500 bp length. Then, the PacBio RS-II sequencing independently generated 598.1 Mb of sequence data (69-fold coverage). After extensive read assembly and correction among these sequencings with the help of sequence data from Roche 454 pyrosequencing, eight contigs were obtained (Fig. 2). Contigs 1 and 2 harbored a 3.1-kb homologous sequence with the right end of pSLA2-L and -M, locating them at both ends of the chromosome. Opposite boundaries in contigs 1 and 2 harbored downstream of 5S-23S rRNA-encoding genes (rDNAs) (gray and blue arrows, respectively, in Fig. 2-ii). Five contigs, 3, 4, 6, 7, and 8, contained downstream of 5S-23S rDNAs and upstream of 16S rDNA (red arrows in Fig. 2-ii) at both ends. Contig 5 harbored 16S rDNA at both ends, indicating the presence of seven rRNA operons (16S-23S-5S) on the *S. rochei* chromosome. Unlinked contig gaps were filled by conventional PCR amplification using KOD-plus Neo DNA polymerase and some group-specific 16S and 5S rRNA primer sets (Table S1). Seven amplified PCR fragments (ca. 6 kb) covering each rRNA operon were sequenced and the connectivities of eight contigs were confirmed. The final assembled sequence revealed an 8,364,802-bp linear chromosome with a G + C content of 71.7% and 7,568 predicted coding DNA sequences (CDSs). Since we previously reported the nucleotide sequences of the three linear plasmids, pSLA2-L (210,614 bp)<sup>10</sup>, -M (113,464 bp)<sup>12</sup>, and -S (17,526 bp)<sup>14</sup>, the gross genomic sequence of strain 7434AN4 has now been determined to be 8,706,406 bp.



**Figure 3.** Terminal inverted repeat (TIR) of the *S. rochei* chromosome. **(a)** Nucleotide sequence comparison of inside ends of the TIR regions. Identical sequences are indicated by asterisks. Ch-L, chromosome left region; Ch-R, chromosome right region. **(b)** Southern hybridization analysis of the TIR regions. PCR fragment harboring nt 52,079–55,053 was used as a DNA probe to distinguish the TIR boundary.  $\lambda$ -DNA digested with *Hind*III was used as a DNA size marker. Lane M,  $\lambda$ /*Hind*III marker; lane 1, *S. rochei* 7434AN4 total DNA digested with *Bam*HI; lane 2, *S. rochei* 7434AN4 total DNA digested with *Xho*I. **(c)** Restriction map of the left and right TIR regions of the *S. rochei* 7434AN4 chromosome. TIR regions of the chromosome are shown by thick black lines. Terminal proteins attached to the 5'-ends are indicated by filled circles. Some important restriction sites are indicated. Bg, *Bgl*II; Ba, *Bam*HI; Xh, *Xho*I; Ec, *Eco*RI; Kp, *Kpn*I.

The linear topology of the chromosome was further supported by pulsed-field gel electrophoresis (PFGE) analysis of *Ase*I and *Dra*I digest (Fig. S2). All of the predicted *Ase*I fragments, namely six fragments larger than 610 kb, three fragments ranging 285–450 kb, and five fragments below 225 kb were detected, except for the largest fragment *Ase*I-A (~2.25 Mb). Southern hybridization analysis of PFGE fragments probed by PCR fragments containing a flanking *Ase*I site agreed with the calculated *Ase*I map (Fig. S2). All of the predicted *Dra*I fragments, four larger than 945 kb, three ranging 285–450 kb, and three below 225 kb, were detected (Fig. S2).

**General features of the linear chromosome of *S. rochei* 7434AN4.** The general features of the *S. rochei* 7434AN4 chromosome are summarized in Tables 1 and 2. In addition, features including distribution of rDNA operons, tRNAs, BGCs, and CDSs according to direction of transcription (+ strand, upper line; – strand, lower line) as well as GC-skew diagram (Fig. 2) were generated by DNA plotter software (<https://www.sanger.ac.uk/science/tools/dnaplotter>)<sup>15</sup>. Linear chromosomes and linear plasmids of *Streptomyces* generally contain TIRs at both ends<sup>1</sup>. Previous Southern blot analysis with the pSLA2-L end probe indicated that the size of the TIRs of the 7434AN4 chromosome is shorter than 70 kb<sup>14</sup>. The present study revealed that the 7434AN4 chromosome has 53,892-bp TIRs (Fig. 3a). The inside ends of the TIRs were analyzed by Southern hybridization using a 3.0-kb PCR fragment (nt 52,079–55,053) as a probe (Fig. 3b). When 7434AN4 total DNA was digested with *Bam*HI, two expected signals appeared at 3.5 kb (left-region of chromosome = ch-L) and 4.6 kb (right-region of chromosome = ch-R). When digested with *Xho*I, two signals were observed at 5.9 kb (ch-L) and 8.6 kb (ch-R). These results coincide with the obtained sequence data. The length of TIR varies among *Streptomyces* species, generally from tens to hundreds kilobases<sup>16</sup>. Ubiquitous presence of relatively long TIRs at both ends of *Streptomyces* linear replicons led to the idea that they might function to maintain a linear topology<sup>17</sup>. TIRs potentially provide a suitable location for homologous recombination, when one TIR is lost by terminal deletion. If recombination occurs inside of TIRs, it regenerates intact TIRs and if does outside of TIRs, it results in arm replacement<sup>18,19</sup>, both of which could recover intact termini. However, exceptionally short TIRs were found in *S. hygroscopicus* 5008 (14 bp)<sup>20</sup> and in *S. avermitilis* (49 bp)<sup>6</sup> (Table 1).

Telomere-associated protein and terminal protein coded by *tap* and *tpg*, respectively, are necessary for terminal replication of *Streptomyces* linear replicons<sup>13</sup>. However, only a truncated *tpg* homolog (SRO\_0241, 136 aa) without a counterpart of *tap* was found on the *S. rochei* chromosome. This gene is identical to the truncated *tpg* homolog in contig 95 reported previously by our group<sup>14</sup> and shows a short but high homology only to the central part of *tpg*; the gene product of SRO\_0241 (136 aa) shows 84% (29/35) and 69% (24/35) identity to that of *tpgR1* of pSLA2-L (184 aa) and *tpgRM* of pSLA2-M (184 aa), respectively. This result suggests that the original *tpg* gene on the chromosome might have been truncated at both 5'- and 3'-sides to generate a fusion gene SRO\_0241 with no function. Thus, the absence of a *tap-tpg* gene pair on the chromosome of strain 7434AN4 has now been confirmed by genome sequencing. In addition, our hypothesis that the lack of *tap-tpg* is rescued in *S. rochei* 7434AN4 by introducing pSLA2-L and -M<sup>14</sup>, has also been proved. In this connection, the smallest plasmid pSLA2-S (17,526 bp) lacks a *tap-tpg* gene pair. It is reasonable that we could not obtain a mutant carrying only pSLA2-S from *S. rochei* 7434AN4, although all other mutants carrying possible combinations of three linear plasmids were obtained<sup>9</sup>.

The *S. rochei* chromosome contained seven rRNA operons (Fig. 2 and Table S2) and 67 tRNA genes (from 43 families) (Table S3). The replication origin *oriC* of *Streptomyces* is generally located between *dnaA* and *dnaN*, and contains at least 19 *dnaA* box-like sequences<sup>21</sup>. A putative *oriC* of the *S. rochei* 7434AN4 chromosome was located at nt. 4,097,668–4,098,732 about 80 kb from the center toward the left end, which also contains 19 *dnaA*-like boxes (Figs S2 and 2).

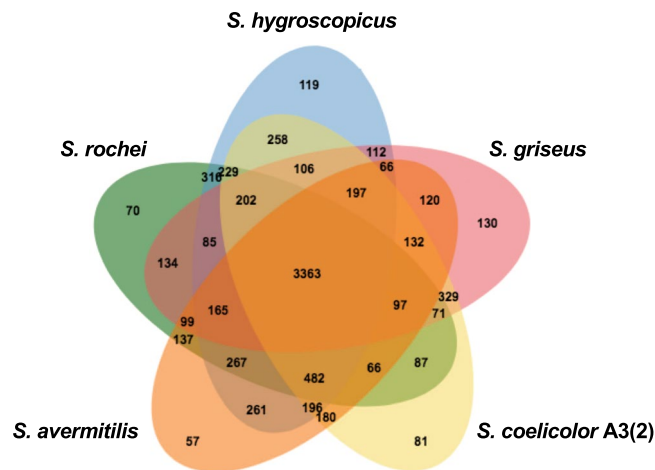
TTA codons are rare in *Streptomyces* species due to their high G + C content (typically more than 70%); for example, TTA-bearing genes comprised 1.7 and 3.4% of the *S. coelicolor* and *S. avermitilis* genomes, respectively<sup>22</sup>. On the *S. rochei* 7434AN4 chromosome, 225 CDSs (2.9%) contain TTA codons (Table S4) (SRO\_0031 and SRO\_7538 are duplicates since they are in the TIR regions of the chromosome). Distribution of TTA-bearing CDSs in five *Streptomyces* genomes, *S. coelicolor*, *S. avermitilis*, *S. griseus*, *S. hygrosopicus*, *S. scabies*, and *S. rochei*, were analyzed by reciprocal BLAST-P search (Table S5). Among 225 TTA-bearing CDSs on the *S. rochei* chromosome, 182 (81%) are specific for *S. rochei* and 30 (13%) are shared with the closely related species *S. hygrosopicus*, suggesting that TTA-bearing CDSs are species-specific in *Streptomyces*. Since TTA, one of the six leucine codons, is rare in streptomycetes, *bldA*, a gene for UUA-specific tRNA, has a crucial role in morphological differentiation and antibiotic production<sup>23–25</sup>. Among 225 TTA-bearing CDSs on the *S. rochei* chromosome, 17 CDSs (yellow boxes in Table S4) are involved in secondary metabolite biosynthesis, including polyketide synthases and non-ribosomal peptide synthetases. In the pentamycin biosynthetic gene cluster (BGC) (Fig. S4), five of 12 CDSs (*pemA1*, *pemA2*, *pemA5*, *pemC*, and *pemR*) contain a TTA codon. On the contrary, the BGC for filipin (=14-deoxo-pentamycin) in *S. avermitilis* contains only one TTA-bearing CDS (*pteR*; a homolog of *pemR*), suggesting that pentamycin production in *S. rochei* is strictly controlled under *bldA*-dependent regulon.

**Comparative analysis of the *S. rochei* chromosome with the genomes of *S. coelicolor* A3(2), *S. avermitilis*, *S. griseus*, and *S. hygrosopicus*.** Most *Streptomyces* species, including *S. coelicolor* A3(2), *S. avermitilis*, and *S. griseus*, have six *rrn* operons (16S–23S–5S), however, *S. rochei* 7434AN4 has seven *rrn* operons (Table 1). Unusual number of *rrn* operons were also reported for *S. albus* J1074 (seven operons)<sup>26</sup> and for *S. xiamenensis* 318 (five operons)<sup>27</sup>. As shown in Fig. S3, six *rrn* operons (*rrnA*, *rrnC*, *rrnD*, *rrnE*, *rrnF*, and *rrnG*) in *S. rochei* are located between the highly conserved ORFs; for example, *rrnA* operon between putative beta-lactamase (SRO\_2104; upstream of 16S) and putative aminotransferase (SRO\_2103; downstream of 5S), and *rrnG* between phosphoenolpyruvate-dependent sugar phosphotransferase (SRO\_6100; upstream of 16S) and CDP-alcohol phosphatidyltransferase (SRO\_6101; upstream of 5S). The boundary regions around the seventh *rrn* operon, designated as *rrnB* in *S. rochei* (SRO\_2782 at upstream of 16S, and SRO\_2781 at upstream of 5S) are apparently different from those in *S. albus* J1074.

We then compared all CDSs on the chromosome of five *Streptomyces* strains, *S. coelicolor* A3(2), *S. avermitilis*, *S. griseus*, *S. hygrosopicus*, and *S. rochei*, using two independent *in silico* analyses; (1) orthologous clustering analysis by OrthoVenn Analysis Software (<http://www.bioinfogenome.net/OrthoVenn/>)<sup>28</sup> and (2) pair-wise genome alignments by GenomeMatcher, a graphical interface for comparative genomics<sup>29</sup>. In ortholog clustering analysis, the CDSs on the *S. rochei* chromosome were classified into 5,370 clusters, among which 3,363 orthologs (44.4%) were shared in 5 strains (Fig. 4). In pair-wise genome alignments, the *S. rochei* chromosome contains a highly conserved core region (around nt 1.31–6.83 Mb) compared with the four reference strains (Fig. S5). However, a large genomic inversion was detected in *S. rochei*. When compared with *S. avermitilis*, a 1.54 Mb inversion was observed at 3.34–4.88 Mb region of the *S. rochei* chromosome, which corresponds to the 6.08–4.31 Mb region of the *S. avermitilis* chromosome.

**Biosynthetic gene clusters (BGCs) for secondary metabolites.** We have focused on secondary metabolite biosynthetic machineries and regulatory pathways coded on pSLA2-L<sup>4,11,30–33</sup>. Several mutations including regulatory genes and major biosynthetic pathways coded on pSLA2-L led to activation of “silent” secondary metabolite clusters for pentamycin, citreodiol, *epi*-cireodiol, and KA57-A (Fig. 1)<sup>34–36</sup>. However, none of these biosynthetic gene clusters were found on pSLA2-L, suggesting their presence on the chromosome. Based on the 2ndFind database search (<http://biosyn.nih.gov/jp/2ndFind/>) together with antiSMASH platform<sup>37</sup>, 35 BGCs for secondary metabolites were predicted on the *S. rochei* chromosome. These BGCs were classified into the following groups (Table 3); 8 for polyketides (PKSs), 8 for non-ribosomal peptides (NRPSs), 3 for hybrid PKS/NRPSs, 3 for lantibiotics, 5 for terpenes, 3 for siderophores, 1 for azoxyalkene, 1 for pseudosugar, 1 for butyrolactone, 1 for melanine, and 1 for ectoine (Table 3). The total length of BGCs (588 kb) occupied 7.03% of the *S. rochei* chromosome, which is comparable to other *Streptomyces* strains (594 kb and 6.6% for *S. avermitilis*).

Some chromosome-borne metabolites were obtained by genome mining on *S. rochei*<sup>36</sup>. The *lkcA* mutant (*lkcA*; an NRPS-PKS hybrid gene involved in lankacidin biosynthesis) overproduced three UV-active compounds,



**Figure 4.** Venn diagram of the number of shared and unique genes between *S. rochei* and four other *Streptomyces* strains. OrthoVenn, a web-based application (<http://www.bioinfo-genome.net/OrthoVenn/>) was used in this analysis. Other *Streptomyces* strains used in this analysis are *S. coelicolor* A3(2), *S. avermitilis* MA-4680, *S. griseus* IFO13350, and *S. hygrosopicus* 5008.

pentamycin, citreodiol, and *epi*-citreodiol (Fig. 1)<sup>34</sup>. Comparison of the pentamycin cluster (*pem*) with the filipin (=14-deoxypentamycin) cluster (*pte*) of *S. avermitilis*<sup>38</sup> revealed their high homology (79–92% identities in ORFs) (Fig. S4), except for two additional genes, a P450 monooxygenase gene *SRO\_7222* (*pemI*) and a ferredoxin gene *SRO\_7221* (*pemJ*), in the former. At this stage, we have not yet identified the biosynthetic gene cluster for citreodiol and *epi*-citreodiol (*ctr* cluster). According to the reported feeding experiment<sup>39</sup>, the *ctr* cluster might include a C-methyltransferase (C-MT) gene for introduction of two methyl groups at C-2 and C-6. Two iterative type-I PKSs containing a C-MT domain (*SRO\_6380* and *SRO\_7330*) are potential candidates for biosynthesis of citreodiols, whose gene inactivation is in progress in our laboratory. We isolated an azoxyalkene compound KA57-A (Fig. 1) from a genetically engineered strain KA57, which contains triple mutations on *srrB* (a *tetR*-type receptor gene), *lkcF-KR1* (a ketoreductase domain 1 of *lkcF* for lankacidin biosynthesis), and *lkmE* (a type-II thioesterase gene for lankamycin biosynthesis) coded on pSLA2-L<sup>35</sup>. KA57-A has a unique azoxy group ( $N=N^+-O^-$ ) and its biosynthetic gene (*azx*) cluster was located at nt 2,061,273–2,097,283 of the chromosome by comparison with the BGC for valanimycin<sup>40</sup>. It is noteworthy that production of these four metabolites coded on the chromosome was activated by mutation of the genes coded on pSLA2-L, indicating that the linear plasmid affects not only on a topology of the linear chromosome but also on secondary metabolite production in *S. rochei*.

*Streptomyces* species produce branched-chain fatty acids for both primary and secondary metabolism<sup>41</sup>. Branched-chain amino acids, isoleucine, valine, and leucine, are converted to the corresponding 2-oxoacids, which were then decarboxylated to form 2-methylbutyryl-CoA, isobutyryl-CoA, and isovaleryl-CoA, respectively, by the branched-chain 2-oxoacid dehydrogenase complex (Fig. S6). In the biosynthesis of *Streptomyces* signaling molecules<sup>42</sup>, branched-chain  $\beta$ -ketoacyl-CoA esters ( $C_8$ – $C_{13}$  in length) are condensed with a dihydroxyacetone phosphate unit by specific enzyme such as AfsA (in *S. griseus*), ScbA (in *S. coelicolor*), BarX (in *S. virginiae*), and *SrrX* (in *S. rochei*). In *S. rochei*, the branched-chain fatty acid starter units for SRB1 and SRB2 (Fig. 1) are isobutyrate and (*S*)-2-methylbutyrate, respectively. Furthermore, the macrolide skeleton of lankamycin (Fig. 1) is also derived from one (*S*)-2-methylbutyrate starter unit and six malonyl-CoA extender units. The biosynthetic gene cluster *bkdFGH* for isobutyrate and (*S*)-2-methylbutyrate<sup>43</sup> was also found on the *S. rochei* chromosome (*SRO\_3599*, 3598, and 3597 for *bkdF*, *G* and *H*, respectively) (Fig. S6).

**Other genes.** Comparative analysis of protein families with other four *Streptomyces* genome (*S. coelicolor* A3(2), *S. avermitilis*, *S. griseus*, and *S. hygrosopicus*) (Table S6) revealed that *S. rochei* 7434AN4 has relatively larger proportions of two-component histidine kinase gene homologs (113 vs 58 on average) and ABC transporter-related genes (338 vs 220 on average), reflecting a great extent of signal transduction and material transport. Strain 7434AN4 harbors 36 sigma factors and 21 ECF sigma factors. Homologous gene encoding the principal sigma factor  $\sigma^{hrdB}$ <sup>44</sup> was identified as *SRO\_2011*. Two sporulation-related sigma factors,  $\sigma^{BldN/AdsA}$  for aerial mycelium formation in *S. coelicolor* A3(2)/*S. griseus*<sup>45,46</sup> and  $\sigma^{WhiG}$  for onset of spore formation in *S. coelicolor* A3(2)<sup>47</sup>, were also identified as *SRO\_3261* and *SRO\_2209*, respectively. In addition, other important genes,  $\sigma^R$  for response to oxidative stress in *S. coelicolor* A3(2)<sup>48</sup> and  $\sigma^{shbA}$  for governing  $\sigma^{hrdB}$  in *S. griseus*<sup>49</sup> were also identified as *SRO\_2590* and *SRO\_2981*, respectively.

CRISPR (clustered regularly interspaced short palindromic repeats) is an RNA-dependent immune system widely distributed in bacteria and archaea against infection of foreign genetic elements including phages and plasmids<sup>50</sup>. The CRISPR-associated genes (*cas* genes) were identified from *SRO\_1948* to *SRO\_1955* (Table S7). Flanking this cluster, 19 DNA repeats (CGGTTACCTCCGCCTGCGCGGAGCGGAC; 29 bases) were located upstream at nt. 2,205,703–2,206,838 and 6 repeats were downstream at nt. 2,216,850–2,217,183 (Table S8). All the Cas proteins showed considerable similarity with annotated *cas* gene products in other *Streptomyces* strains. The linear plasmid pSLA2-M has 49 repeat sequences at the right end of the CRISPR cluster (ORF94-ORF101)<sup>12</sup>,

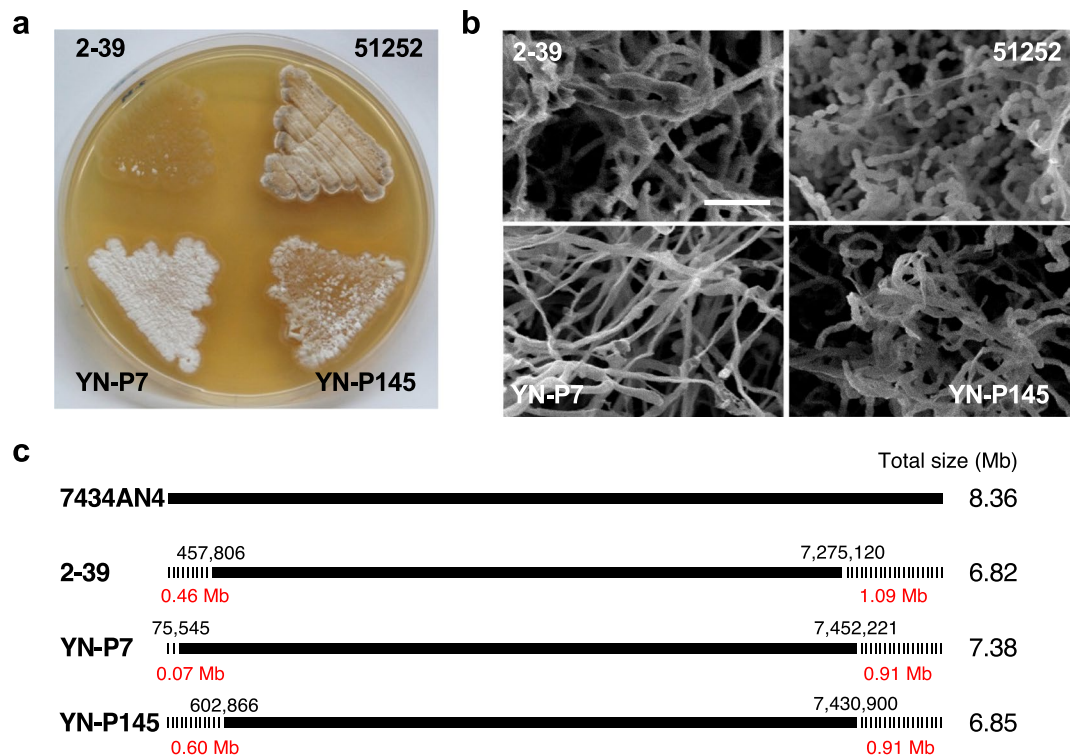


Cluster <sup>a</sup>	Product type	Location	Length (bp)	Possible metabolites
1	PKS-NRPS	SRO_0155-SRO_0157	12,171	
2	Siderophore	SRO_0365-SRO_0370	8,474	
3	Terpene	SRO_0471-SRO_0472	2,153	2-Methylisoborneol
4	PKS-NRPS	SRO_0501-SRO_0517	49,191	
5	PKS	SRO_0730-SRO_0739	15,137	
6	Lantipeptide-NRPS	SRO_0832-SRO_0848	34,433	
7	PKS-NRPS	SRO_1003-SRO_1008	13,813	
8	NRPS-Lantipeptide	SRO_1184-SRO_1189	11,055	
9	Terpene	SRO_1200-SRO_1204	6,711	Hopene
10	NRPS	SRO_1300-SRO_1346	70,044	
11	NRPS	SRO_1551-SRO_1557	21,989	
12	PKS-NRPS	SRO_1613-SRO_1615	11,979	
13	Terpene	SRO_1735-SRO_1737	4,337	Geosmin
14	Azoxyalkene and type I PKS	SRO_1819-SRO_1850	36,010	KA57-A
15	Siderophore	SRO_2036-SRO_2038	4,744	
16	NRPS	SRO_2051-SRO_2053	5,458	
17	NRPS	SRO_2500-SRO_2506	9,881	
18	Terpene	SRO_2585-SRO_2586	2,410	Albaflavenone
19	Butyrolactone	SRO_3382	932	
20	NRPS-Lantipeptide	SRO_3845-SRO_3849	6,875	
21	Type II PKS	SRO_4037-SRO_4044	7,077	Spore pigment
22	Siderophore	SRO_4684-SRO_4688	5,997	
23	Melanin	SRO_4761-SRO_4762	1,446	Melanin
24	Ectoine	SRO_5592-SRO_5595	3,165	5-Hydroxyectoine
25	Type I PKS	SRO_6166-SRO_6168	12,248	
26	Type III PKS	SRO_6250-SRO_6251	2,284	
27	NRPS	SRO_6287-SRO_6291	9,037	
28	NRPS	SRO_6304-SRO_6312	46,390	
29	Type I PKS (Iterative)	SRO_6377-SRO_6381	12,789	
30	Pseudosugar	SRO_6636-SRO_6646	12,173	
31	NRPS	SRO_7209-SRO_7215	44,813	
32	Type I PKS	SRO_7221-SRO_7235	81,482	Pentamycin
33	Type I PKS (Iterative)	SRO_7329-SRO_7331	9,456	
34	NRPS	SRO_7379-SRO_7381	3,029	
35	Terpene	SRO_7452-SRO_7458	8,979	Carotenoid

**Table 3.** List of secondary metabolite biosynthetic gene clusters in the *S. rochei* chromosome. <sup>a</sup>Secondary metabolite gene clusters were predicted by 2<sup>nd</sup> Find and antiSMASH. The range of gene clusters for unknown secondary metabolites was estimated on the basis of putative functions of gene products.

however, whose consensus sequences, 5'-GTGGCGGTGCGCCCTCCGGGGTGACCGAGGATCGCAAC-3' (37 bases), are different from that on the chromosome.

**Analysis of three plasmidless mutants, *S. rochei* 2-39, YN-P7, and YN-P145.** We previously prepared three plasmidless mutants, 2-39, YN-P7, and YN-P145 by protoplast regeneration of *S. rochei* 51252 that contains only pSLA2-L<sup>9,14</sup>. In the case of mutant 2-39, chromosomal deletion at both ends followed by circularization was confirmed by cloning of the fusion junction<sup>14</sup>. Although mutant 2-39 lost around 20% (1.55 Mb) of the chromosome (Fig. 5), seven rRNA operons were still conserved. PCR amplification showed that strains YN-P7 and YN-P145 also keep seven *rrn* operons, suggesting the importance of all *rrn* operons in *S. rochei*. Three mutants exhibited a different phenotype when grown on YM solid medium (Fig. 5); mutants YN-P7 and YN-P145 showed a “white” phenotype, while strain 2-39 a “bald” phenotype (Fig. 5a). To analyze morphological differentiation more precisely, their colonies were observed by scanning electron microscopy (SEM) (Fig. 5b). Chain elongation of aerial mycelium of mutant 2-39 stopped at an early stage, while mutant YN-P7 produced longer but collapsed hyphae. On the other hand, mutant YN-P145 produced partially spiral spore chains although their development is significantly lesser than that in strain 51252. Based on the Illumina sequence data of the three mutants, the bald strain 2-39 was found to suffer a larger chromosomal deletion from the right end (1,090 kb) compared with strains YN-P7 (913 kb) and YN-P145 (934 kb) (Fig. 5c). On the other hand, the deletions at the left chromosomal end of strains 2-39, YN-P7, and YN-P145 were 458 kb, 76 kb, and 603 kb, respectively. Based on the phenotype-genotype correlation in strains 2-39, YN-P7 and YN-P145, we speculate that essential



**Figure 5.** Morphological differentiation of three plasmidless mutants of *S. rochei* and their chromosomal deletion. **(a)** Spore formation of *S. rochei* strains. Strains (2-39, YN-P7, YN-P145, and their parent 51252) were grown on YM agar medium at 28 °C for 5 days. **(b)** Scanning electron microscopy (SEM) of surface grown colonies. **(c)** Chromosomal deletions in mutants 2-39, YN-P7, and YN-P145.

gene(s) responsible for converting vegetative hyphae into aerial hyphae is(are) located at nt 7,275–7,431 kb of the 7434AN4 chromosome. This region harbors 138 ORFs (SRO\_6607-SRO\_6744), among which some gene(s) may be responsible for aerial hyphae formation in *S. rochei*.

## Conclusion

In this study, we have determined the nucleotide sequence of the 8,364,802-bp linear chromosome of *S. rochei* 7434AN4, which in turn revealed the gross genome size (8,706,406 bp) of this strain including the three linear plasmids, pSLA2-L, -M, and -S. General features of the linear chromosome were presented; it carries seven *rrn* operons, 67 tRNA genes, 225 TTA-containing CDSs, and 53.9-kb TIRs at both ends. In particular, the absence of a *tpg-tap* gene pair on the chromosome has proved our hypothesis that the *tpg-tap* pairs of pSLA2-L and/or pSLA2-M function to maintain a linear topology of the chromosome in strain 7434AN4.

*In silico* analysis indicated the presence of 35 secondary metabolites gene clusters on the chromosome, whose functions are not known in most cases. Therefore, we could expect that studies on their functions and regulation, particularly interaction with the regulatory genes coded on pSLA2-L will lead to a discovery of new antibiotics and their improved production.

## Materials and Methods

**Strains, plasmids, oligonucleotides, and culture media.** All the strains, plasmids, and oligonucleotides used in this study were listed in Table S1. YEME liquid medium (0.3% yeast extract, 0.5% peptone, 0.3% malt extract, 1.0% D-glucose, 34% sucrose, 5 mM MgCl<sub>2</sub>, and 0.5% glycine) was used for preparation of total genomic DNA. YM medium (0.4% yeast extract, 1.0% malt extract, and 0.4% D-glucose, pH 7.3) was used for routine cultivation.

**DNA sequencing and assembly.** *S. rochei* 7434AN4 was sequenced using hybrid approach of next-generation sequencing platforms; PacBio RS-II, Illumina GAIIx, and Roche 454 sequencers.

Genomic DNA of strain 7434AN4 was subjected to paired-end sequencing using Illumina GAIIx sequencing system (San Diego, CA, USA) according to the manufacturer's protocol. The 26,119,215 trimmed reads with 217-fold coverage of the whole genome were assembled using ABySS 1.3.7<sup>51</sup>. Illumina read data has been deposited as DRA Accession DRA003131 and DRA003132, Bioproject: PRJDB3565, Biosample: SAMD00027156. Independently, long-read sequencing was performed on PacBio RS-II sequencing system (Pacific Biosciences; Menlo Park, CA, USA). The filtered subreads with 598,149,379 bp in length (69-fold coverage) from PacBio RS-II was then assembled using the Hierarchical Genome Assembly Process (HGAP). The assembly consists of 175 contigs of 8,204,607 bp with an average length of 46,883 bp. Both sequence data was extensively compared and corrected with a help of Roche 454 pyrosequencing to obtain eight contigs (Fig. 2; length of contigs 1-8 were

2,393,094 bp, 1,685,790 bp, 1,286,111 bp, 750,875 bp, 720,184 bp, 567,335 bp, 501,529 bp, and 462,928 bp, respectively. Details for Roche 454 pyrosequencing were described previously<sup>14</sup>. Sequence gaps among eight contigs were then religiously filled and connected by conventional PCR amplification. Complete nucleotide sequence of the *S. rochei* 7434AN4 chromosome has been deposited at DDBJ under Accession number AP018517.

**Sequence annotation and comparative analysis.** Putative coding sequences (CDSs), tRNA-, and rRNA-coding sequences were predicted using Microbial Genome Annotation Pipeline (MiGAP) platform (<https://www.migap.org/>) and FramePlot 2.3.2 (<http://www.nih.gov/~jun/cgi-bin/frameplot.pl>)<sup>52</sup>. Their putative annotation was manually confirmed by a BLASTP program (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>). Main features including distribution of rDNA operons, tRNAs, BGCs, and CDSs according to direction of transcription (+ strand, upper line; – strand, lower line) as well as GC-skew diagram (Fig. 2) were generated by DNA plotter software (<https://www.sanger.ac.uk/science/tools/dnplotter>)<sup>15</sup>. The locus of *oriC* was predicted manually based on the genome information of other *Streptomyces* species. The protein families were clustered with OrthoVenn Analysis Software (<http://www.bioinfogenome.net/OrthoVenn/>), a web platform for comparison and annotation of orthologous gene clusters among multiple species<sup>28</sup>. The comparative analysis of the chromosomes between *S. rochei* 7434AN4 and other *Streptomyces* species was performed using GenomeMatcher software (<http://www.ige.tohoku.ac.jp/joho/gmProject/gmhomeJP.html>)<sup>29</sup> and bl2seq program, which is embedded in the application bundled. Secondary metabolite gene clusters were predicted by either 2ndFind software, a web-based analytical tool (<http://biosyn.nih.gov/2ndFind/>), or antiSMASH 2.0, a web-based analysis platform (<http://antismash.secondarymetabolites.org/>)<sup>37</sup>. CRISPRs were predicted using a CRISPRFinder (<http://crispr.i2bc.paris-saclay.fr/Server/>), an online program.

**DNA manipulation and Southern hybridization.** *Streptomyces* strains were grown in liquid YM medium in Sakaguchi flask at 28 °C for 3 days. DNA manipulation of *Streptomyces* species<sup>53</sup> was carried out according to standard procedure. Genome DNA sample of *S. rochei* 7434AN4 for PFGE was obtained according to the method as described previously<sup>9</sup> with a slight modification. Polymerase chain reaction (PCR) was performed on a GeneAmp G02 Thermal Cycler (Astec Co. Ltd., Fukuoka, Japan) using KOD-plus Neo DNA polymerase (Toyobo, Osaka, Japan) according to the manufacturer's protocol. Southern blot analysis (Figs 3, 5 and S2) was performed as described previously<sup>14</sup>.

**Scanning electron microscopy (SEM).** The surface morphology of *S. rochei* strain 51252 and three plasmidless mutants was observed by scanning electron microscopy (SEM) after growing on YM agar plate for 5 days. For the preparation of specimens, agar plugs were fixed with 1% osmium tetroxide solution for 12 h, and then dehydrated by lyophilization. The resulting specimens were coated with platinum (2 nm) and observed by a Jeol JSM-5900 Scanning Electron Microscope.

## References

- Lin, Y.-S., Kieser, H. M., Hopwood, D. A. & Chen, C. W. The chromosomal DNA of *Streptomyces lividans* 66 is linear. *Mol Microbiol.* **10**, 923–933 (1993).
- Chen, C. W. *et al.* The linear chromosomes of *Streptomyces*: Structure and dynamics. *Actinomycetol.* **8**, 103–112 (1994).
- Volff, J. N. & Altenbuchner, J. Genetic instability of the *Streptomyces* chromosome. *Mol Microbiol.* **27**, 239–246 (1998).
- Kinashi, H. Antibiotic production, linear plasmids and linear chromosomes in *Streptomyces*. *Actinomycetol.* **22**, 20–29 (2008).
- Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature.* **417**, 141–147 (2002).
- Ikeda, H. *et al.* Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol.* **21**, 526–531 (2003).
- Ohnishi, Y. *et al.* Genome Sequence of the Streptomycin-Producing Microorganism *Streptomyces griseus* IFO 13350. *J Bacteriol.* **190**, 4050–4060 (2008).
- Hayakawa, T., Tanaka, T., Sakaguchi, K., Otake, N. & Yonehara, H. A linear plasmid-like DNA in *Streptomyces* sp. producing lankacidin group antibiotics. *J Gen Appl Microbiol.* **25**, 255–260 (1979).
- Kinashi, H., Mori, E., Hatani, A. & Nimi, O. Isolation and characterization of linear plasmids from lankacidin-producing *Streptomyces* species. *J Antibiot.* **47**, 1447–1455 (1994).
- Mochizuki, S. *et al.* The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. *Mol Microbiol.* **48**, 1501–1510 (2003).
- Arakawa, K., Tsuda, N., Taniguchi, A. & Kinashi, H. The butenolide signaling molecules SRB1 and SRB2 induce lankacidin and lankamycin production in *Streptomyces rochei*. *Chembiochem.* **13**, 1447–1457 (2012).
- Yang, Y. *et al.* pSLA2-M of *Streptomyces rochei* is a composite linear plasmid characterized by self-defense genes and homology with pSLA2-L. *Biosci Biotechnol Biochem.* **75**, 1147–1153 (2011).
- Bao, K. & Cohen, S. N. Recruitment of terminal protein to the ends of *Streptomyces* linear plasmids and chromosomes by a novel telomere-binding protein essential for linear DNA replication. *Genes Dev.* **17**, 774–785 (2003).
- Nindita, Y. *et al.* The *tap-tpg* gene pair on the linear plasmid functions to maintain a linear topology of the chromosome in *Streptomyces rochei*. *Mol Microbiol.* **95**, 846–858 (2015).
- Carver, T., Thomson, N., Bleasby, A., Berriman, M. & Parkhill, J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics (Oxford, England)* **25**, 119–20 (2009).
- Chen, C. W., Huang, C. H., Lee, H. H., Tsai, H. H. & Kirby, R. Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes. *Trends Genet.* **18**, 522–529 (2002).
- Qin, Z., Stanley, N. & Cohen, S. N. Survival mechanisms for *Streptomyces* linear replicons after telomere damage. *Mol Microbiol.* **45**, 785–794 (2002).
- Fischer, G., Wenner, T., Decaris, B. & Leblond, P. Chromosomal arm replacement generates a high level of intraspecific polymorphism in the terminal inverted repeats of the linear chromosomal DNA of *Streptomyces ambofaciens*. *Proc. Natl. Acad. Sci. USA* **95**, 14296–14301 (1998).
- Uchida, T., Miyawaki, M. & Kinashi, H. Chromosomal arm replacement in *Streptomyces griseus*. *J Bacteriol.* **185**, 1120–1124 (2003).
- Wu, H. *et al.* Genomic and transcriptomic insights into the thermo-regulated biosynthesis of validamycin in *Streptomyces hygroscopicus* 5008. *BMC Genomics.* **13**, 337 (2012).
- Jakimowicz, D. *et al.* Structural elements of the *Streptomyces oriC* region and their interactions with the DnaA protein. *Microbiology.* **144**, 1281–1290 (1998).

22. Chandra, G. & Chater, K. F. Evolutionary flux of potentially *bldA*-dependent *Streptomyces* genes containing the rare leucine codon TTA. *Anton Leeuw*. **94**, 111–126 (2008).
23. Leskiw, B. K., Lawlor, E. J., Fernandez-Abalos, J. M. & Chater, K. F. TTA codons in some genes prevent their expression in a class of developmental, antibiotic-negative, *Streptomyces* mutants. *Proc Natl Acad Sci USA* **88**, 2461–2465 (1991).
24. Takano, E. *et al.* A rare leucine codon in *adpA* is implicated in the morphological defect of *bldA* mutants of *Streptomyces coelicolor*. *Mol Microbiol*. **50**, 475–486 (2003).
25. White, J. & Bibb, M. *bldA* dependence of undecylprodigiosin production in *Streptomyces coelicolor* A3(2) involves a pathway-specific regulatory cascade. *J Bacteriol*. **179**, 627–633 (1997).
26. Zaburannyi, N., Rabyk, M., Ostash, B., Fedorenko, V. & Luzhetskyy, A. Insights into naturally minimised *Streptomyces albus* J1074 genome. *BMC Genomics*. **15**, 97 (2014).
27. Xu, M.-J. *et al.* Deciphering the streamlined genome of *Streptomyces xiamenensis* 318 as the producer of the anti-fibrotic drug candidate xiamenmycin. *Sci Rep*. **6**, 18977 (2016).
28. Wang, Y., Coleman-Derr, D., Chen, G. & Gu, Y. Q. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucl Acids Res*. **43**, W78–84 (2015).
29. Ohtsubo, Y., Ikeda-Ohtsubo, W., Nagata, Y. & Tsuda, M. GenomeMatcher: A graphical user interface for DNA sequence comparison. *BMC Bioinform*. **9**, 376 (2008).
30. Arakawa, K., Sugino, F., Kodama, K., Ishii, T. & Kinashi, H. Cyclization mechanism for the synthesis of macrocyclic antibiotic lankacidin in *Streptomyces rochei*. *Chem Biol*. **12**, 249–256 (2005).
31. Arakawa, K., Kodama, K., Tatsuno, S., Ide, S. & Kinashi, H. Analysis of the loading and hydroxylation steps in lankamycin biosynthesis in *Streptomyces rochei*. *Antimicrob Agents Chemother*. **50**, 1946–1952 (2006).
32. Arakawa, K., Mochizuki, S., Yamada, K., Noma, T. & Kinashi, H.  $\gamma$ -Butyrolactone autoregulator-receptor system involved in lankacidin and lankamycin production and morphological differentiation in *Streptomyces rochei*. *Microbiology*. **153**, 1817–1827 (2007).
33. Yamamoto, S., He, Y., Arakawa, K. & Kinashi, H. Gamma-butyrolactone-dependent expression of the *Streptomyces* antibiotic regulatory protein gene *srrY* plays a central role in the regulatory cascade leading to lankacidin and lankamycin production in *Streptomyces rochei*. *J Bacteriol*. **190**, 1308–1316 (2008).
34. Cao, Z., Yoshida, R., Kinashi, H. & Arakawa, K. Blockage of the early step of lankacidin biosynthesis caused a large production of pentamycin, citreodiol and epi-citreodiol in *Streptomyces rochei*. *J Antibiot*. **68**, 328–333 (2015).
35. Kunitake, H., Hiramatsu, T., Kinashi, H. & Arakawa, K. Isolation and biosynthesis of an azoxyalkene compound produced by a multiple gene disruptant of *Streptomyces rochei*. *ChemBioChem*. **16**, 2237–2243 (2015).
36. Arakawa, K. Manipulation of metabolic pathway controlled by signaling molecules, inducers of antibiotic production, for genome mining in *Streptomyces* spp. *Antonie Leeuw*. **111**, 743–751 (2018).
37. Blin, K. *et al.* AntiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucl Acids Res*. **41**, W204–12 (2013).
38. Ikeda, H., Shin-ya, K. & Ōmura, S. Genome mining of the *Streptomyces avermitilis* genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. *J Ind Microbiol Biotechnol*. **41**, 233–250 (2014).
39. Shizuri, Y., Nishiyama, S., Imai, D. & Yamamura, S. Isolation and stereostructures of citreoviral, citreodiol, and epicitreodiol. *Tetrahedron Lett*. **25**, 4771–4774 (1984).
40. Garg, R. P., Ma, Y., Hoyt, J. C. & Parry, R. J. Molecular characterization and analysis of the biosynthetic gene cluster for the azoxy antibiotic valanimycin. *Mol Microbiol*. **46**, 505–517 (2002).
41. Han, L., Lobo, S. & Reynolds, K. A. Characterization of beta-ketoacyl-acyl carrier protein synthase III from *Streptomyces glaucescens* and its role in initiation of fatty acid biosynthesis. *J Bacteriol*. **180**, 4481–4486 (1998).
42. Takano, E.  $\gamma$ -Butyrolactones: *Streptomyces* signaling molecules regulating antibiotic production and differentiation. *Curr Opin Microbiol*. **9**, 287–294 (2006).
43. Denoya, C. D. *et al.* A second branched-chain alpha-keto acid dehydrogenase gene cluster (*bkdFGH*) from *Streptomyces avermitilis*: its relationship to avermectin biosynthesis and the construction of a *bkdF* mutant suitable for the production of novel antiparasitic avermectins. *J Bacteriol*. **177**, 3504–3511 (1995).
44. Buttner, M. J., Chater, K. F. & Bibb, M. J. Cloning, disruption, and transcriptional analysis of three RNA polymerase sigma factor genes of *Streptomyces coelicolor* A3(2). *J Bacteriol*. **172**, 3367–3378 (1990).
45. Bibb, M. J., Molle, V. & Buttner, M. J.  $\sigma^{bldN}$ , an extracytoplasmic function RNA polymerase sigma factor required for aerial mycelium formation in *Streptomyces coelicolor* A3(2). *J Bacteriol*. **182**, 4606–4616 (2000).
46. Yamazaki, H., Ohnishi, Y. & Horinouchi, S. An A-factor-dependent extracytoplasmic function sigma factor ( $\sigma^{AdsA}$ ) that is essential for morphological development in *Streptomyces griseus*. *J Bacteriol*. **82**, 4596–4605 (2000).
47. Chater, K. F. *et al.* The developmental fate of *S. coelicolor* hyphae depends upon a gene product homologous with the motility sigma factor of *B. subtilis*. *Cell*. **59**, 133–143 (1989).
48. Paget, M. S., Kang, J. G., Roe, J. H. & Buttner, M. J.  $\sigma^R$ , an RNA polymerase sigma factor that modulates expression of the thioredoxin system in response to oxidative stress in *Streptomyces coelicolor* A3(2). *EMBO J*. **17**, 5776–5782 (1998).
49. Otani, H., Higo, A., Nanamiya, H., Horinouchi, S. & Ohnishi, Y. An alternative sigma factor governs the principal sigma factor in *Streptomyces griseus*. *Mol Microbiol*. **87**, 1223–1236 (2013).
50. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol*. **1**, e60 (2005).
51. Simpson, J. T. *et al.* ABySS: a parallel assembler for short read sequence data. *Genome Res*. **19**, 1117–1123 (2009).
52. Ishikawa, J. & Hotta, K. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G+C content. *FEMS Microbiol Lett*. **174**, 251–253 (1999).
53. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *Practical Streptomyces Genetics*. (John Innes Foundation, Norwich, 2000).

## Acknowledgements

We thank to Dr. D. Kajiya and Mrs. T. Amimoto (N-BARD, Hiroshima University) for measurement of the high resolution mass spectra, and Mrs. K. Koike (N-BARD, Hiroshima University) for SEM analysis. This work was supported by Grants-in-Aid for Scientific Research on Innovative Areas (23108515, 25108718 and 17H05446 to K.A.) from Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT), Grants-in-Aid for Scientific Research (B) (16H04917 to K.A.) from the Japan Society for the Promotion of Science (JSPS), and the Sasakawa Scientific Research Grant from the Japan Science Society to Y.N. This work was partly supported by a JSPS A3 Foresight Program. A.A.F. and R.M. were supported by the Indonesia Endowment Fund for Education (LPDP). Sequencing analysis using an Illumina GAIIx sequencer was supported by the Grant in Aid for Scientific Research on Innovative Areas (22108010 to J.I.) from MEXT.

### Author Contributions

Y.N., Z.C., A.A.F., K.I., H.K. and K.A. designed the experiments, Y.N., Z.C., A.A.F., Y.M. and K.I. performed the experiments, Y.N., A.A.F., A.T., Y.M., R.M., Y.Y., K.I. and K.A. analyzed the data, Y.S., H.Y., M.T., A.L., J.I., M.K. and T.S. performed the next-generation sequencing and extensive assembly, and Y.N., K.I., H.K. and K.A. wrote the manuscript with input from all of the authors. All the authors approved the final version of the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-47406-y>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019