

Supplementary information for
Analysis of a genomic island housing genes for DNA S-modification
system in *Streptomyces lividans* 66 and its counterparts in other
distantly related bacteria

Xinyi He, Hong-Yu Ou, Qing Yu, Xiufen Zhou, Jun Wu, Jingdan Liang, Wei Zhang,
Kumar Rajakumar and Zixin Deng

Supplementary Results

Detailed bioinformatics analysis of the modules of SLG

Module I, comprising two widely separated terminal parts of a linearized form [Ia (*slg01–slg05*) and Ib (*slg82–slg85*)], likely originated as one continuous region in the circular form before integration via its internal *att* site (Fig. 1 & Fig. 2A). This module contains flanking direct repeats (DRs) and genes encoding hypothetical proteins for integration and excision of the island (Fig. 1A). The interval between *attL* and its P4-type cognate integrase was 105 bp, similar to the sizes of three sequenced *Streptomyces* phages (17 bp in ϕ C31: NC_001978; 34 bp in ϕ BT1: NC_004664; 158 bp in phage VWB: NC_005345). The *slg02* CDS was predicted to be a transposase pseudogene as it possessed an apparent 5'-truncation of 330 nucleotides compared with its multiple homologues in *S. coelicolor* A3(2). In addition, we found 14-bp inverted repeats (IRs) flanking *slg02*, 12-bp IRs flanking *slg02–slg03* and 139-bp DRs flanking *slg02–slg03*. (Fig. S1). This suggested that module Ia may have undergone multiple recombination events and thus possibly has presented a recombination hotspot.

Module II is comprised of *slg06–slg25* which encodes several predicted proteins but exhibiting weak homology to those documented in *S. coelicolor* A3(2) and *S.*

avermitilis MA-4680 (Table S1). Notably, the majority of these hypothetical proteins show significant similarity to proteins involved in a nucleotide-related metabolic pathway (green in Fig. 1B; Table S1) including Slg14 (thymidylate synthase), Slg15 (carbamoyltransferase), Slg18 (S - adenosylmethionine - dependent methyltransferase), Slg20 (phosphoglycerate mutase), Slg22 (nucleotide – diphosphate - sugar epimerases), Slg23 (adenosylhomocysteinase) and Slg25 (CMP – 2 – keto – 3 - deoxyoctulosonic acid synthetase). Furthermore, based on intergenic spacing and the unidirectional transcription, 15 adjacent genes (*slg11* - *slg25*) seem to be organised as four operons (*slg12* – *slg14*, *slg15* – *slg16*, *slg17* – *slg20*, and *slg22* – *slg25*).

Module III, which is comprised of a large fragment (*slg45-71*) appears to be syntenic (95% full-length identity) to a region in *S. coelicolor* A3(2) (*SCO3509-34*, highlighted in red in Fig. 1B). We used the *slg60-63* region as a probe to hybridize against the genome of *S. lividans* ZX1 (SLG⁻ strain) but found no positive signal (data not shown), suggesting that module III may have resulted from chromosomal rearrangement rather than DNA duplication in *S. lividans*. In addition, a potential DNA swap of a 1.3-kb *S. coelicolor* A3(2)-like fragment (*SCO3521-22*) with a 4.8-kb G+C valley (*slg56-59*) resulted in acquisition of resistance to ϕ HAU3 by *S. lividans* 66 (Zhou *et al.*, 1994b). Homologues of the *slg58-59* couplet are also found in *Pseudomonas syringae* pv. tomato DC3000, *Mesorhizobium loti* R7A, *Rhodopseudomonas palustris* CGA009 and *Pasteurella multocida* genomes (Fig. S5). Interestingly, the couplet is located in an IS element in *P. syringae* pv. tomato DC3000 (Alfano *et al.*, 2000). Nevertheless, no homologue of *SCO3521* and *SCO3522* was detected in *S. lividans* 66 by Southern hybridization (data not shown). Therefore, gene loss was likely the evolutionary force in shaping module III of SLG (Finan, 2002).

Module IV consists of *slg72* - *81* and harbors a well-characterized *dnd* cluster (*slg76* - *80*), which confers an unusual DNA modification on *S. lividans* 66 (Zhou *et al.*, 2005) and a putative SAM-dependent methyltransferase (*slg81*). These six genes

exhibit synteny to a putative GI-borne region in *S. avermitilis* MA-4680. The previously reported 8.2-kb *dnd* cluster contains five genes involved in incorporation of sulphur or a sulphur-containing substance into the *S. lividans* genome. This *in vivo* DNA modification renders DNA sensitive to *in vitro* oxidative double-strand cleavage and degradation during electrophoresis (Zhou *et al.*, 1988).

Illustration and confirmation of the excision process of pJTU1515 from pJTU1514

This was first achieved by introducing pJTU1514 from *E. coli* ET12567/pJTU1514 into *S. lividans* ZX1 by conjugal transfer at 28°C to generate strain HXY7, followed by a round of selective growth on apramycin-containing agar for 72 hours at 42°C to obtain an apramycin-resistant strain.

Comparison of enzymatic pattern between pJTU1515 and pJTU1514 confirmed that pJTU1515 was derived from pJTU1514. The plasmid DNA prepared from a representative apramycin-resistant clone designated HXY10 was subsequently introduced by electroporation into *E. coli* DH10B. Six randomly selected apramycin resistant colonies were found to harbour a plasmid of 5.7 kb, substantially smaller than pJTU1514 (11.3 kb). When digested with *Apa*I, *Nhe*I or *Ssp*I, all six 5.7-kb plasmids showed the same fragment pattern, and were designated pJTU1515 (Fig. 3B). A comparison of the restriction fragment pattern between pJTU1514 and pJTU1515 revealed that all of the detectable restriction sites flanking outward to *attL* and *attR* on pJTU1514 (Fig. 3A, dotted line) were deleted in pJTU1515, whereas those flanking inward to *attL* and *attR* (Fig. 3A, solid line) remained intact. Thus pJTU1515 was putatively derived from pJTU1514.

Sequence of PCR product targeting putative fusion junction confirmed that pJTU1515 had been formed by circularization mediated by *attL/attR* junction as a specific 0.8-kb fragment could be amplified by PCR using P-pJTU1515F and P-pJTU1515R as primers and pJTU1515 as template (Lane 4 in Fig. 3D), which was subsequently confirmed by sequencing of the resultant 0.8-kb PCR product. Using the same PCR primers and HXY10 total DNA as template, a weak 0.8-kb band could also

be detected (Lane 2 and 3 in Fig. 3D), suggesting that the non-replicative mini-island excised site-specifically from the HXY10 chromosome at low frequency.

References

- Alfano, J.R., Charkowski, A.O., Deng, W.L., Badel, J.L., Petnicki-Ocwieja, T., van Dijk, K., and Collmer, A. (2000) The *Pseudomonas syringae* Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants. *Proc Natl Acad Sci U S A* **97**: 4856-4861.
- Finan, T.M. (2002) Evolving insights: symbiosis islands and horizontal gene transfer. *J Bacteriol* **184**: 2855-2856.
- Kumar, S., Tamura, K., and Nei, M. (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**: 150-163.
- Zhou, X., Deng, Z., Firmin, J.L., Hopwood, D.A., and Kieser, T. (1988) Site-specific degradation of *Streptomyces lividans* DNA during electrophoresis in buffers contaminated with ferrous iron. *Nucleic Acids Res* **16**: 4341-4352.
- Zhou, X., Deng, Z., Hopwood, D.A., and Kieser, T. (1994b) *Streptomyces lividans* 66 contains a gene for phage resistance which is similar to the phage lambda ea59 endonuclease gene. *Mol Microbiol* **12**: 789-797.
- Zhou, X., He, X., Liang, J., Li, A., Xu, T., Kieser, T., Helmann, J.D., and Deng, Z. (2005) A novel DNA modification by sulphur. *Mol Microbiol* **57**: 1428-1438.

Table S1. Characteristics of the predicted *orf* found on the 93 kb SLG island in *S. lividans* 66

No.	Size (aa)	Potential protein homologue(s)	Source organism	Amino acid identity (No. identical/total) [percentage identity]	BlastP E-value	Accession No. (Best match)
01	424	Phage integrase contains domain (177-411) matches Pfam entry phage_integrase (PF00589)with Evalues of 0.00068	<i>Nocardioides</i> sp. JS614	142/423 [33%]	1e-47	ZP_00656461
02	173	Truncated IS4 transposase contains domain (2-153) matches Pfam entry (PF01609)with Evalues of 2.5e-16	<i>Frankia</i> sp. EAN1pec	105/167 [62%]	3e-42	ZP_00569867
03	361	Hypothetical protein contains potential domain (53-73) matches Pfam entry G-protein gamma like domains (PF00631) with Evalues of 0.13				
04	89	Hypothetical protein contains potential domain (18-45) matches Pfam entry Oxidored_q4 (PF00507)with Evalues of 0.86 and Contains transmembrane helix (5-27)				
05	155	Hypothetical protein contains potential domain (107-136) matches Pfam entry RTBV_P46 (PF06216) with Evalues of 0.61	<i>Rhodococcus</i> sp. RHA1	38/105 [36%]	9e-06	YP_703166
06	614	Hypothetical protein contains potential domain (65-96) matches Pfam entry (PF08238) Sel1 repeat with Evalues of 0.46				
07	241	Hypothetical protein; COG1595, RpoE, DNA-directed RNA polymerase specialized sigma subunit				
08	132	Hypothetical protein, contains potential domain (1-88) matches Pfam entry Ribosomal Proteins L2, C-terminal domain (PF03947) with Evalues of 0.077				
09	191	CutA, divalent ion tolerance protein, contains domain (88-189) matches Pfam entry CutA1 (PF03091) with Evalues of 4.8e-30	<i>Shewanella baltica</i> OS155	47/107 [43%]	9e-19	ZP_00580758

10	413	Twin-arginine translocation pathway signal protein, contains domain (8-69) matches Pfam entry HTH_3 (P01381) with Evalues of 1.5e-05; and domain (93-242) matches Pfam-B (PB164262)	<i>Frankia</i> sp. EAN1pec	79/225 [35%]	8e-13	ZP_00572634
11	201	Hypothetical protein contains potential domain (165-190) matches Pfam entry Tenui_NCP (PF04876) with Evalues of 0.013				
12	230	HAD-superfamily hydrolase, subfamily IA, contains domain (20-197) matches Pfam entry Hydrolase (PF00702) with Evalues of 2.1e-21	Uncultured bacterium 313	50/181 [27%]	3e-13	AAR37567
13	650	Hypothetical polysaccharide biosynthesis protein contains potential domain (349-590) matches Pfam entry Capsule_synth (PF05159) with Evalues of 0.075	<i>Bacillus thuringiensis</i> serovar israelensis ATCC 35646	49/191 [25%]	9e-04	ZP_00741533
14	308	Predicted thymidylate synthase contains domain (10-256) matches to Pfam entry Thymidylat_synt (PF00303) with Evalues of 4.5e-08	<i>Aurantimonas</i> sp. SI85-9A1	100/302 [33%]	1e-34	ZP_01228433
15	590	Predicted carbamoyl transferase contains domain (132-524) matches to Pfam entry CmcH_NodU (PF02543) with Evalues 5.2e-70	<i>Hahella chejuensis</i> KCTC 2396	213/634 [33%]	1e-76	YP_431553
16	354	Hypothetical protein				
17	183	DNA topology modulation protein; COG0563, Adk, Adenylate kinase and related kinases	<i>Bacillus halodurans</i> C-125	64/172 [37%]	5e-28	BAB04733
18	317	Methyltransferase contains domain (64-233) matches to Pfam entry Ubie_methyltran (PF01209) with Evalues of 5e-05	<i>Myxococcus xanthus</i> DK1622	56/162 [34%]	2e-11	ABF91632
19	255	Predicted phosphotransferase contains potential domain (2-210) matches to Pfam entry APH (PF01636) with Evalues of 0.022				
20	218	Predicted Fructose-2,6-bisphosphatase phosphoglycerate mutase	<i>Frankia</i> sp. Cc13	100/208 [48%]	3e-46	YP_481446
21	107	Hypothetical protein contains potential domain (2-27) matches to Pfam entry APH (PF01636) with Evalues of 4e-06	<i>S. coelicolor</i> A3(2)	38/104 [36%]	1e-07	NP_639624
22	332	NDP-sugar epimerase; COG1028, FabG, dehydrogenases	<i>Geobacillus kaustophilus</i> HTA426	142/310 [45%]	4e-60	BAD75124

23	375	S-adenosyl-L-homocysteine hydrolase; cd00401, AdoHcyase, <i>Frankia</i> sp. CcI3	192/374 [51%]	6e-98	YP_483169	
		S-adenosyl-L-homocysteine hydrolase				
24	186	NUDIX hydrolase; COG0494, MutT, NTP pyrophosphohydrolases including oxidative damage repair enzymes	<i>Frankia</i> sp. EAN1pec	71/171 [41%]	1e-25	ZP_00570366
25	254	3-deoxy-D-manno-octulosonate cytidyltransferase; COG1212, KdsB, <i>Halothermothrix orenii</i> H 168		102/241 [42%]	7e-41	ZP_01189425
		CMP-2-keto-3-deoxyoctulosonic acid synthetase				
26	368	Hypothetical protein contains five transmembrane helixes spanning residues: 24-46, 67-89 104-126, 143-165 and 180-202, respectively.	<i>Streptomyces griseus</i>	133/346 [38%]	2e-53	AAQ08941
27	209	HAD-superfamily hydrolase subfamily IA; COG1011: predicted hydrolase	<i>Thermobifida fusca</i> YX	88/222 [39%]	8e-27	AAZ56517
28	108	Hypothetical protein				
29	339	Hypothetical protein contains domain (22-166) matches to Pfam-B (PB075407) and potential domain (133-148) matches Pfam entry C_tripleX (PF02363) with Evalues of 0.53	<i>Clostridium thermocellum</i> ATCC 27405	39/120 [32%]	1e-09	ZP_00510568
30	250	Hypothetical protein contains domain (158-237) matches to Pfam-B (PB000826); Potential domain (66-140) matches to Pfam entry DUF1205 (PF06722) with Evalues of 0.81	<i>Streptomyces kanamyceticus</i>	73/148 [49%]	1e-25	BAE95455
31	184	Predicted transcriptional protein contains domain (15-85) matches to Pfam entry Whib (PF02467) with Evalues of 1.5e-10	<i>S. coelicolor</i> A3(2)	66/134 [49%]	2e-22	CAB92584
32	355	Hypothetical protein contains domain (21-143) matches to Pfam-B (PB059328) and potential domain (304-324) matches to Pfam entry peroxidase (PF00141) with Evalues of 0.28 ; three transmembrane helixes spanning residues 47-69,79-98 and	<i>Streptomyces kanamyceticus</i>	115/330 [34%]	3e-29	BAE95458

		119-141, respectively						
33	74	Hypothetical protein						
34	294	Hypothetical protein contains potential domain (234-257) matches to Pfam entry LysR_substrate (PF03466) with Evalues of 0.27						
35	592	Hypothetical protein contains domain (7-279) matches to Pfam entry Relaxase (PF03432) with Evalues of 0.00052	<i>S. coelicolor</i> A3(2)	311/586 [53%]	1e-98	NP_630985		
36	107	Hypothetical protein similar to SCO6917	<i>S. coelicolor</i> A3(2)	27/71 [38%]	1e-04	NP_630984		
37	146	Hypothetical protein contains a potential domain (28-40) matches to Pfam entry Glyco_hydro_25 (PF01183) with Evalues of 0.26						
38	169	Hypothetical protein contains potential domain (41-61) matches to Pfam entry Sep15_SelM (PF08806) with Evalues of 0.87 and domain (129-137) matches to Pfam entry InvH (Pf04741)						
39	156	Hypothetical protein						
40	70	Hypothetical protein similar to SCO6914						
41	129	Hypothetical protein contains domain (66-129) of unknown function matches to Pfam entry SPDY (PF03771) with Evalues of 5.9e-10						
42	150	Hypothetical protein contains potential domain (70-90) of Formyl transferase matches to Pfam entry (PF00551) with Evalues of 0.19						
43	299	Hypothetical protein contains domain (34-93) of unknown function just in streptomyces matches to Pfam entry SPDY (PF03771) with Evalues of 6.1e-13; another potential domain (173-200) match to Pfam entry SPDY (PF03771)with Evalues of 0.23	<i>S. coelicolor</i> A3(2)	85/289 [29%]	2e-09	NP_630970		
44	282	SCO6894, transcriptional regulators; COG1733, predicted transcriptional regulators	<i>S. coelicolor</i> A3(2)	66/208 [31%]	2e-09	NP_630964		
45	250	SCO3534, possible large ATP-binding protein, contains Prosite match to PS00017 ATP/GTP-binding site motif A (P-loop) and a probable coiled-coil region around 600-650aa	<i>S. coelicolor</i> A3(2)	140/156 [89%]	5e-53	NP_627734		

46	221	SCO3533, hypothetical protein	<i>S. coelicolor</i> A3(2)	192/221 [86%]	1e-85	NP_627733
47	158	SCO3532, hypothetical protein	<i>S. coelicolor</i> A3(2)	119/129 [92%]	3e-45	NP_627732
48	259	SCO3531, hypothetical protein contains domain (5-222) matches to Pfam-B (PB107524); another potential domain (156-207) matches to Pfam entry DUF551(PF04448) with Evalues of 0.31	<i>S. coelicolor</i> A3(2)	236/259 [91%]	7e-117	NP_627731
49	199	SCO3530, hypothetical protein contains domain (99-176) matches to Pfam-B (PB029607); another domain (26-60) matches to Pfam entry_SOR/SNZ (Pf01680) with Evalues of 0.36	<i>S. coelicolor</i> A3(2)	189/199 [94%]	1e-86	NP_627730
50	316	SCO3529, hypothetical proteins contains potential domain (1-119) matches to Pfam entry Auxin (PF05564) with Evalues of 0.058	<i>S. coelicolor</i> A3(2)	149/173 [86%]	2e-60	NP_627729
51	529	SCO3528, DnaB-like helicase contains domains (1-98, 191-241) match to Pfam DnaB (PF00772) with Evalues of 9.5e-05 and 0.00018, respectively	<i>S. coelicolor</i> A3(2)	318/344 [92%]	4e-131	NP_627728
52	66	Hypothetical protein identical to SCO3527	<i>S. coelicolor</i> A3(2)	53/64 [82%]	4e-08	NP_627727
53	422	SCO3526, cd01126, TraG_VirD4, The TraG/TraD/VirD4 family are bacterial conjugation proteins	<i>S. coelicolor</i> A3(2)	418/422 [99%]	0.0	NP_627726
54	146	SCO3525, possible transmembrane protein, TMhelixes spanning residues (17-39,77-99); translocase binding subunit	<i>S. coelicolor</i> A3(2)	123/131 [93%]	9e-50	NP_627725
55	493	Hypothetical protein identical to SCO3524 and SCO3523 contains potential domain (5-170) matches to Pfam-B.	<i>S. coelicolor</i> A3(2)	181/185 [97%]	7e-84	NP_627724
			<i>S. coelicolor</i> A3(2)	226/239 [94%]	2e-92	NP_627723
56	156	Hypothetical protein similar to SAV329 contains domain (24-149) matches to Pfam-B; another potential domain (5-36) matches to Pfam entry HEPPP_synt_1 (PF07307) with Evalues of 0.095	<i>S. avermitilis</i> MA-4680	103/128 [80%]	8e-35	NP_821503
57	300	Tn574I family transposase similar to SAV254 contains domain (126-281) matches to Pfam entry transposaes_11 (PB01609) with Evalues of 7.4e-21 and another potential domain matches to Pfam-B (PB000143), respectively	<i>S. avermitilis</i> MA-4680	171/275 [62%]	1e-84	NP_821428
58	299	Possible EA31 gene protein contains domains (73-116, 120-262) matches to Pfam-B	<i>Rhodopseudomonas</i>	81/191 [42%]	5e-38	CAE28008

		(PB023654) and Pfam-B (PB028282), respectively	<i>palustris</i> CGA009			
59	497	Phage φHAU3 resistance protein	<i>Streptomyces lividans</i>	341/343 [99%]	0.0	CAA54681
60	248	SCO3520, similar to GTP-binding elongation factor contains potential domain (6-81)Pfam-B (P130729)	<i>S. coelicolor</i> A3(2)	188/198 [94%]	5e-91	NP_627720
61	187	Hypothetical protein identical to SCO3519 contains domain (21-156) matches to Pfam-B (PB169412)	<i>S. coelicolor</i> A3(2)	148/165 [89%]	3e-51	NP_627719
62	211	Hypothetical protein identical to SCO3518	<i>S. coelicolor</i> A3(2)	196/211 [92%]	2e-89	NP_627718
63	489	Hypothetical protein contains potential domains (2-121, 122-189, and 190-487) match to Pfam-B_104063, Pfam-B_001029 and Pfam-B_012716, respectively. Another two potential domains (124-298, 318-325) match to Pfam Entry AAA (PF00004) and Pfam entry TraH_2 (PF06871) with Evalues of 0.58 and 0.47, respectively	<i>S. coelicolor</i> A3(2)	482/489 [98%]	0.0	NP_627717
64	138	Hypothetical protein identical to SCO3516 contains domain (1-137) matches to Pfam-B (PB129756)	<i>S. coelicolor</i> A3(2)	132/138 [95%]	7e-54	NP_627716
65	477	Transmembrane protein contains two TMhelixes spanning residues (10-32, 39-58); It contains domains (57-337, 338-477) match to Pfam-B059687 and Pfam-B 134786; another two potential domains (210-381, 405-426) match Pfam entry DUF1849 (PF08904) and DUF1956 (PF09209) with Evalues of 0.69 and 0.78, respectively	<i>S. coelicolor</i> A3(2)	453/462 [98%]	0.0	NP_627715
66	458	Transmembrane protein contains six TMhelixes spanning residues (61-83, 95-117, 151-173, 180-198, 208-230 and 250-272); This protein contains domains (1-283 and 404-434) match to Pfam-B_165229 and Pfam-B_1156, respectively. Contains potential domains (78-292, 99-312) match to Pfam entry 7tm_1 (PF00001), Pfam entry MASE1 (PF05231) with Evalues of 0.8 and 0.51, respectively	<i>S. coelicolor</i> A3(2)	442/458 [96%]	0.0	NP_627714
67	200	Possible secreted protein contains domain (38-198) match to Pfam-B (PF104702)	<i>S. coelicolor</i> A3(2)	171/188 [90%]	1e-70	NP_627713
68	102	Transmembrane protein contains two TMhelixes spanning residues (36-58, 73-95); This protein contains potential domain (3-102) matches to Pfam-B_182646; another	<i>S. coelicolor</i> A3(2)	100/102 [98%]	9e-22	NP_627712

		potential domain (85-97) matches to Pfam entry TRAP-gamma (PF07074) with Evalues of 0.83					
69	330	Hypothetical protein contains domain (217-328) matches to Pfam entry NLPC_P60(PF00877) with Evalues of 1.7e-28; This protein contains another three potential domains match to Pfam-B_147800, Pfam-B_009444 and Pfam-B_170236, respectively	<i>S. coelicolor</i> A3(2)	316/330 [95%]	7e-137	NP_627711	
70	248	Possible DNA methylase contains domain (22-230) matches to Pfam entry N6_N4Mtase(PF01555) with Evalues of 1.6e-22	<i>S. coelicolor</i> A3(2)	244/248 [98%]	2e-132	NP_627710	
71	289	COG1637, Predicted nuclease of the RecB family contains domains (3-138, 190-287) matches to Pfam-B_031133 and Pfam-B_012579, respectively	<i>S. coelicolor</i> A3(2)	276/289 [95%]	2e-136	NP_627709	
72	692	Hypothetical protein	<i>Anabaena variabilis</i> ATCC 29413	143/583 [24%]	1e-34	ABA25018	
73	576	SAV3726, pfam02518, HATPase_c, histidine kinase-, DNA gyrase B, and HSP90-like ATPase	<i>S. avermitilis</i> MA-4680	300/577 [51%]	6e-156	NP_824903	
74	133	Hypothetical protein contains potential domain (21-56) matches to Pfam entry GIY-YIG (PF01541)					
75	156	Hypothetical protein contains domain (33-42) matches to Pfam entry CPSase_L_D3 (PF02787) with Evalues of 0.63	<i>S. avermitilis</i> MA-4680	46/131 [35%]	4e-10	NP_824904	
76*	126	DndE; SAV2929, conserved hypothetical protein	<i>S. avermitilis</i> MA-4680	72/116 [62%]	5e-40	NP_824105	
77*	628	DndD; SAV2930, SbcC, putative ABC transporter ATP-binding protein	<i>S. avermitilis</i> MA-4680	337/678 [49%]	3e-144	NP_824106	
78*	498	DndC; SAV2931, 3'-phosphoadenosine-5'-phosphosulfate sulfotransferase	<i>S. avermitilis</i> MA-4680	347/515 [67%]	0.0	NP_824107	
79*	312	DndB; SAV2932, hypothetical protein	<i>S. avermitilis</i> MA-4680	219/312 [70%]	3e-121	NP_824108	

80*	397	DndA; SAV2933, IscS, putative L-cysteine desulfurase (transpersulfidase) (NifS protein homolog)	<i>S. avermitilis</i>	272/384 [70%]	1e-148	NP_824109
			MA-4680			
81*	478	SAV2934, putative SAM-dependent methyltransferase	<i>S. avermitilis</i>	338/476 [71%]	0.0	NP_824110
			MA-4680			
82	134	Hypothetical protein				
83	80	Hypothetical protein				
84	253	Hypothetical protein				
85	210	Hypothetical protein contains domain (135-147) matches to Pfam entry tRNA 2e (PF02091) with Evalues of 0.07				

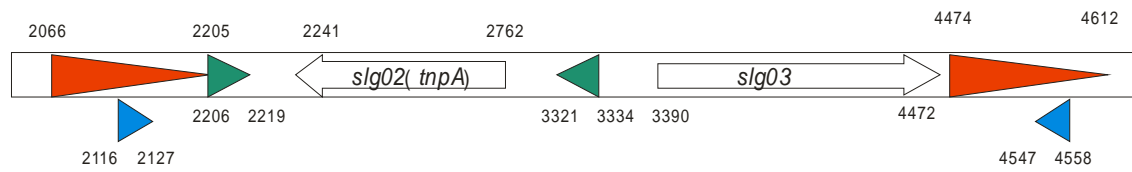
* The 8026 bp *S. lividans* 66 sequence that coded for *slg76* to *slg81* had been previously deposited under the GenBank accession number AAZ29041.

δ The protein-coding region of *slg55* contains two adjacent individual orfs (SCO3523 and SCO3524) in the genome annotation of the *Streptomyces coelicolor* A3(2)... (http://www.sanger.ac.uk/Projects/S_coelicolor/)

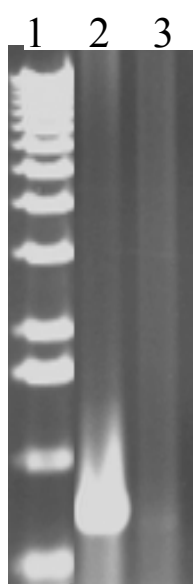
Table S2. Oligonucleotide primers used in this study

Primers	Forward (F)	Reverse (R)
Lp1	ATGGCGTTCACGGACTGG	TGCCCTGCGACACCTCTAC
Rp1	GGCTTTACACTTTATGCTTCCG	CATCCACGACTGGGTCTACG
Rp2	GTCGGGCACTCAGGAACAC	
P-pJTU1515	CCTGCGACACCTCTACG	GCTTTACACTTTATGCTTCC
attR	GTGCCTCCGSSCTGCT	CCACGACTGGGTCTACG
LC2	GGCGAAGACGGCGAACA	TGCGGCTGGTGAAGGAGTAG
P21	TGAAAGGCGATGAAGGTAGG	CAGAAGGCTCGTTTAGACAGG
Lter	CTTCCTCTTCCCTCCCTCCTCA	GCAGCAGTTCGGGCTTGGT
M1	CGCTGAGTTGCTGGACATTG	CCTCGCTGAGAACCACCTTG
M2	CTGCGTGAGCATCGGGAAGA	GGCTGGCTGAGGGTGAAGAA
M3	AAGAATGCGGTAGTGGAGCG	TCGGACTTCGTGGAGTGGAT
M4	TCAAAGGTGGATTGGGTTCG	ATGGCGGAGGTGATGATGTG
Rter	ATGGACCACGAGCGTTTGTT	GCACCTTCAGGCTGTAGGGA
aac(3)IV-T	CCGACTGGACCTTCCTTCT	GACACGATGCCAACACGAC
aadA-T	TCGGCAGCGACATCCTTC	CGTCATCGAGCGCCATCT
PSCO3507	CCCCTGCTGCTATTCGG	TCCCTTGCTCGTCTTTTCG
P1	GCCACCACTTCAAGAACTCTGT	TGTTATCCGCTCACAATTCCAC
P2	ACCTTTGCCATCAACTCGG	GGCGGTAATACGGTTATCCAC

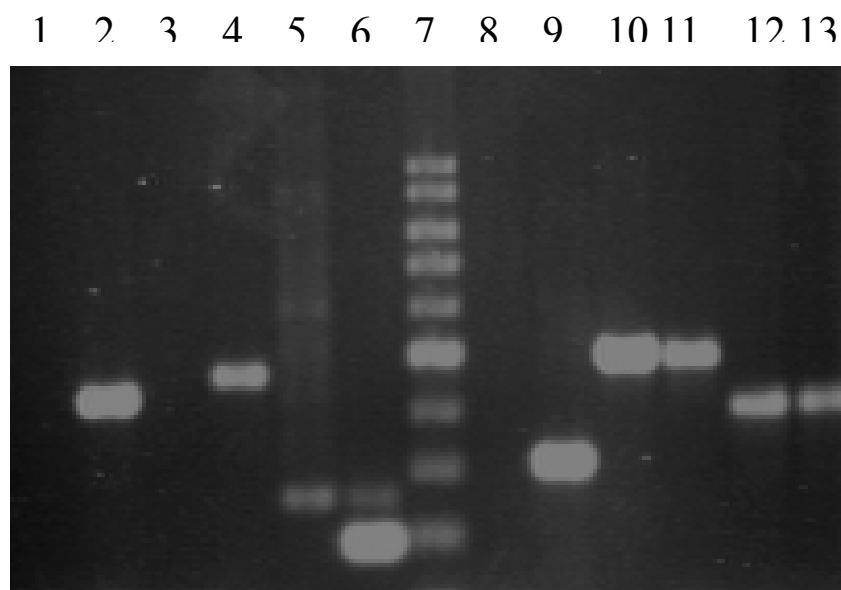
Supplementary Figures



Supplementary Fig. S1. Schematic diagram supporting the region flanking the *slg02* and *slg03* as a recombination hotspot. colored triangle present pairs of IR or DR, while the digits represents the start and end sites of IR, DR and *orf*.



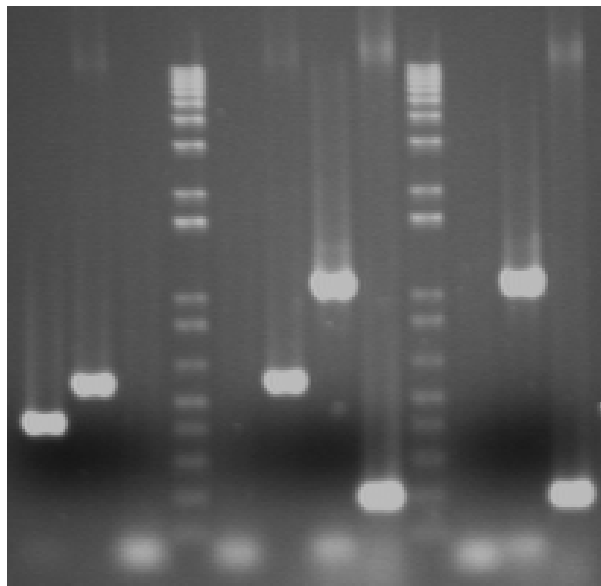
A



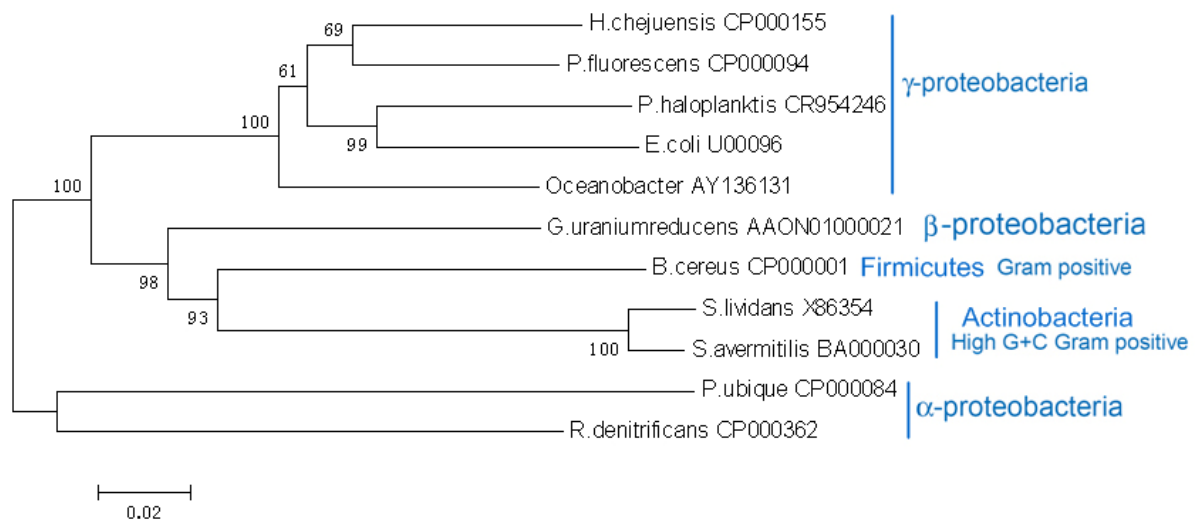
B

Supplementary Fig. S2. Confirmation of the SLG excision coupled with the loss of resistance to streptomycin from the chromosome of *S. lividans* HXY16 with the MNNG induction.

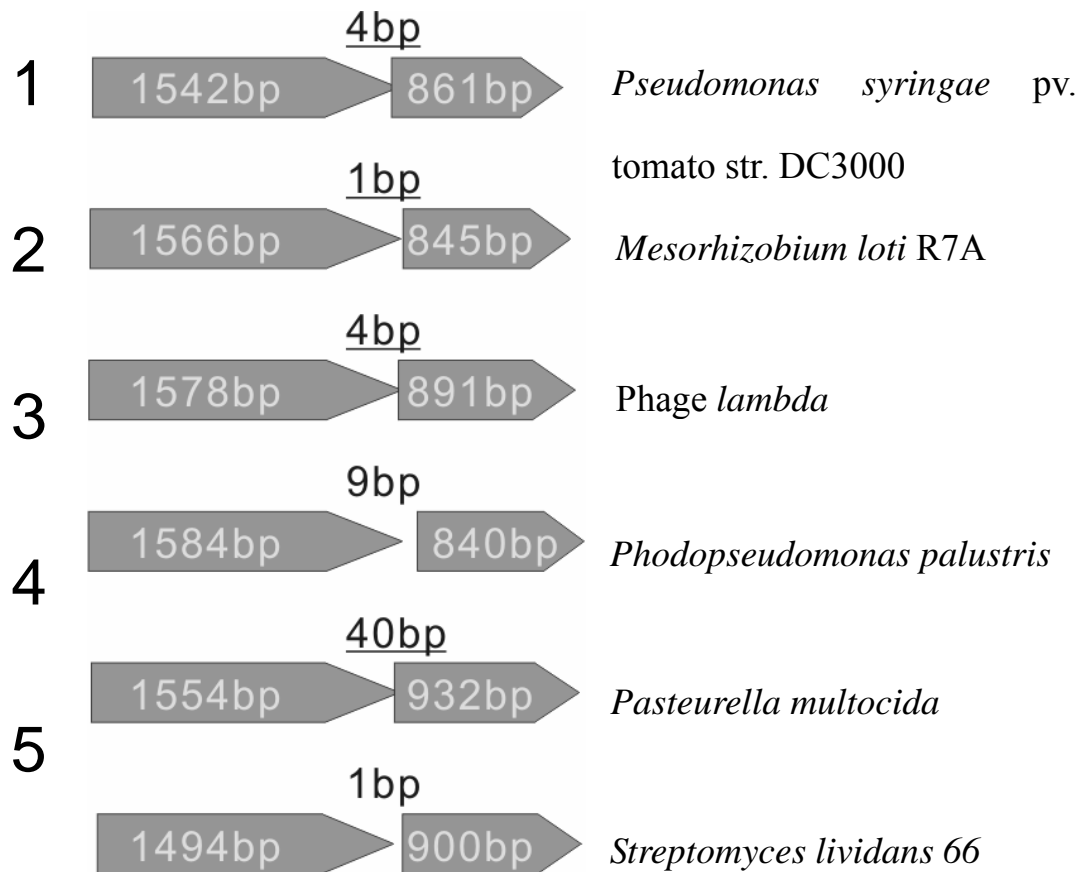
1 2 3 4 5 6 7 8 9 10 11 12



Supplementary Fig. S3. Proof that the exconjugants acquired the resistant marker other than the SLG from the donor *S. lividans* HXY1.



Supplementary Fig. S4. Inferred phylogenetic relationship of the 11 bacterial strains carrying a *dnd* cluster (or their close relatives) on the basis of 16S rDNA sequences. The tree was constructed using a MEGA 3.1 alignment (Kumar *et al.*, 2004) and the Neighbour Joining method.



Supplementary Fig. S5. Schematic diagram representing the conservative organization of ϕ HAU3^r-ea31 loci in bacteria and phages, implying its possible organization as an operon and a functional unit.