Molecular Microbiology (2007) **65**(4), 1034–1048
doi:10.1111/j.1365-2958.2007.05846.x
First published online 19 July 2007

# 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,[1] Hong-Yu Ou,[1] Qing Yu,[1] Xiufen Zhou,[1] Jun Wu,[1] Jingdan Liang,[1] Wei Zhang,[2] Kumar Rajakumar[3,4] and Zixin Deng[1]***

[1]*Laboratory of Microbial Metabolism and School of Life Science and Biotechnology, Shanghai Jiaotong University, Shanghai 200030, China.*
[2]*Division of Biology, Illinois Institute of Technology, Chicago, IL 60616, USA.*
[3]*Department of Infection, Immunity and Inflammation, Leicester Medical School, University of Leicester, Leicester LE1 9HN, UK.*
[4]*Department of Clinical Microbiology, University Hospitals of Leicester NHS Trust, Leicester LE1 5WW, UK.*

## Summary

**The complete sequence (92 770 bp) of a genomic island (GI) named SLG from *Streptomyces lividans* 66, encoding a novel DNA S-modification system (*dnd*), was determined. Its overall G+C content was 67.8%, lower than those of three sequenced *Streptomyces* genomes. Among 85 predicted open reading frames (ORFs) in SLG, 22 ORFs showed little homology with previously known proteins. SLG displays a mosaic structure composed of four modules, indicative of multiple recombination events in its formation. Spontaneous excision and circularization of SLG was observed, and the excision rate appeared to be induced at least fivefold by MNNG exposure. Using constructed mini-islands of SLG, we demonstrated that Slg01, a P4-like integrase, was sufficient to promote SLG integration, excision and circularization. Eleven counterpart *dnd* clusters, which also mapped to GIs in 10 chromosomes and a plasmid, were found in taxonomically unrelated bacterial species from various geographic niches. Additionally, *c.* 10% of actinomycetes were found to possess a *dnd* cluster in a survey involving 74 strains. Comparison of *dnd* clusters in the 12 bacteria strongly suggests that these *dnd*-bearing elements might have evolved from a common ancestor similar to plasmid-originated chromosome II of *Pseudoalteromonas haloplanktis* TAC125.**

Accepted 19 June, 2007. *For correspondence. E-mail zxdeng@sjtu.edu.cn; Tel. (+86) 21 6293 3404; Fax (+86) 21 6293 2418.



## Introduction

*Streptomyces* species are soil-dwelling filamentous bacteria that produce most known natural antibiotics as well as many other secondary metabolites and secreted enzymes of economic and industrial importance (Hopwood, 2007). Considerable phenotypic variation is commonly observed in many genera of actinomycetes. This is predominantly attributed to high levels of chromosomal instability caused by homologous or illegitimate recombination events that frequently result in deletions, insertions, amplifications and/or rearrangements (Volff and Altenbuchner, 1998; Gunes *et al*., 1999). Horizontal gene transfer events contribute to further genome instability via the integration of unstable alien mobile genetic elements into the *Streptomyces* linear chromosome. Examples of acquired elements include insertion sequences (ISs), transposons, prophages, conjugative plasmids and genomic islands (GIs) (Bentley *et al*., 2002; Ikeda *et al*., 2003; Choulet *et al*., 2006). The 'mobilome' (mobile genome) (Ou *et al*., 2005) of an individual organism has been hypothesized to reflect the bacterium's lifestyle, pathogenicity, adaptation to particular ecological niches and evolutionary history (Dobrindt *et al*., 2004). Interestingly, a subset of GIs associated with secondary metabolism has recently been identified in many bacterial species and these are proposed as important players in the moulding of natural product biosynthesis (Dobrindt *et al*., 2004; Piel *et al*., 2004). Most integrases encoded by GIs recognize the 3′ termini of tRNA genes as integration hotspots (Ou *et al*., 2006), whereas some integrases target other conserved loci such as *thdF* (in *Salmonella* SGI1) (Doublet *et al*., 2005) and *bacA* (in *Streptomyces scabies* pathogenicity island) (Kers *et al*., 2005). Island-borne integrases or 'cross-talking' homologues typically mediate the process of site-specific integration and excision (Hochhut *et al*., 2006). In some cases, additional excisionases or other auxiliary factors are required (Doublet *et al*., 2005). In addition, some GIs are readily transmissible under standard laboratory conditions and are consequently thought to have been recently acquired. Examples of mobile GIs include the 500 kb

symbiosis island in *Mesorizobium loti* (Sullivan and Ronson, 1998), SXT GI in *Vibrio cholerae* (Hochhut and Waldor, 1999) and *clc* GI of *Pseudomonas* sp. strain B-13 (Sentchilo *et al*., 2003a,b). In contrast, many chromosomally integrated GIs seem to have lost transmissibility. Several of these 'immobile' structures exhibit highly mosaic content, indicating the likely occurrence of multiple recombination events (Hsiao *et al*., 2003; Mantri and Williams, 2004).

The sizes of linear *Streptomyces* chromosomes range from the 8.7 Mb chromosome in *S. coelicolor* A3(2) (Bentley *et al*., 2002) to the 10.1 Mb replicon in *S. scabies* (http://www.sanger.ac.uk/Projects/S_scabies/). These linear *Streptomyces* chromosomes appear to be compartmentalized into 'core' and 'arm' regions. The central core region contains the essential genes, whereas the arms carry conditionally adaptive genes and species-specific DNA (Bentley *et al*., 2002). Despite the high level of synteny between the core regions of actinomycete chromosomes, several GIs have been found in these regions (Bentley *et al*., 2002; Ikeda *et al*., 2003; Choulet *et al*., 2006). For example, 12 large insertions were identified in the core region of the *S. coelicolor* chromosome following comparison of its genome with that of a close relative, *S. ambofaciens* (Choulet *et al*., 2006). Most of these were probably acquired recently because they contain discernible mobility-associated features. Application of the Islander algorithm, which detects GIs adjacent to tRNA sites bounded by direct repeats (DRs) and containing an integrase gene homologue (Mantri and Williams, 2004), led to the identification of five and three islands in the chromosomes of *S. coelicolor* A3(2) and *S. avermitilis* MA-4680 respectively. A large (325–660 kb), mobile pathogenicity island was reported to be conserved among three plant pathogenic *Streptomyces* species (*S. acidoscabies*, *S. scabies* and *S. turgidiscabies*), which encodes a pathogenicity determinant, the phytotoxin thaxtomin (Kers *et al*., 2005).

Another important example is the *c.* 90 kb *S. lividans* SLG, found inserted into the 3′-end of the chromosomal *murA1* gene, and absent from *S. coelicolor* A3(2) (Zhou *et al*., 2004). The SLG island contains a phage ϕHAU3 resistance gene (*ϕHAU3′*) (Zhou *et al*., 1994a) and a 8.2 kb five-gene cluster involved in site-specific incorporation of sulphur (Zhou *et al*., 2005) into *S. lividans* DNA (Dyson and Evans, 1998, Liang *et al*., 2007). We named this SLG-mediated *in vivo* DNA modification the Dnd phenotype. Although the role of this modification is not yet known, modified DNA is sensitive to *in vitro* oxidative double-strand cleavage and degradation during normal and pulsed-field gel electrophoresis (PFGE) (Zhou *et al*., 1988).

In order to better understand the function and evolution of SLG in *Streptomyces,* we sequenced and characterized SLG in *S. lividans* 66. We found that SLG has a mosaic composition and shows features of multiple horizontal acquisition events. We demonstrated that an SLG-encoded P4-type integrase mediated site-specific integration, excision and circularization of SLG, while native SLG itself is apparently non-transmissible. The *dnd* cluster is widely distributed among distantly related bacterial species. Most strikingly, the identified *dnd* clusters appear to be borne on large integrated elements, supporting the existence of a diverse 'family' of GIs that carry this novel and intriguing DNA modification system.

## Results

### *SLG, a mosaic-like region characteristic of an unusual GI in* S. lividans

The 93 kb DNA is thought to be inserted into the *S. lividans* 66 chromosome at the 3′ terminus of *murA1* with a recognizable flanking DR that matched the 15 bp 3′-end of *murA1* (Zhou *et al*., 2004). Analysis of the SLG sequence revealed an intact integrase gene homologue (*slg01*) and a truncated transposase gene (*slg02*) in the *murA1*-proximal end (Zhou *et al*., 2004). Eighty-five putative protein-coding genes, *slg01–slg85*, were predicted with FramePlot beta 3.0 (Ishikawa and Hotta, 1999), and found to occupy 76% of the island sequence (Fig. 1B). The results of a BLASTP (Schaffer *et al*., 2001), PFAM (Finn *et al*., 2006) and TMHMM (Krogh *et al*., 2001) searches for predicted proteins and its putative functions are shown in Table S1. Twenty-two of them showed little homology with proteins in databases. SLG has a markedly lower G+C content (67.8%) than the genome of *S. coelicolor* A3(2) (72.1%), a very close relative of *S. lividans* 66 (Kieser *et al*., 1992; Leblond *et al*., 1993), suggesting that SLG could have been acquired laterally from elsewhere. A distinct dinucleotide bias of the island region compared with the complete sequence of *S. coelicolor* A3(2) also strongly supported a foreign origin of SLG (Table 1).

For the convenience of description and discussion, the SLG island was divided into four putative modules defined on the basis of G+C content, clustered functional organization, sequence homology and/or AT-rich valleys (Fig. 1): module I (11.9 kb, 64.3% G+C) harbours a putative recombination hotspot (Fig. 1A; Fig. S1), module II (20.4 kb, 65.4% G+C) putatively encodes a nucleotide-related metabolic pathway (green in Fig. 1B; Table S1), module III (46.2 kb, 68.8% G+C) seemed to cover a large syntenic DNA fragment (*slg45–71*) with 95% full-length nucleotide identity to a region in *S. coelicolor* A3(2) (*SCO3509–34,* highlighted in red in Fig. 1B) and module IV (14.2 kb, 65.1% G+C) encompasses the *dnd* gene cluster (blue in Fig. 1B; Table S1). (Detailed bioinformatics analysis of these four modules is provided in the *Supplementary material*.)

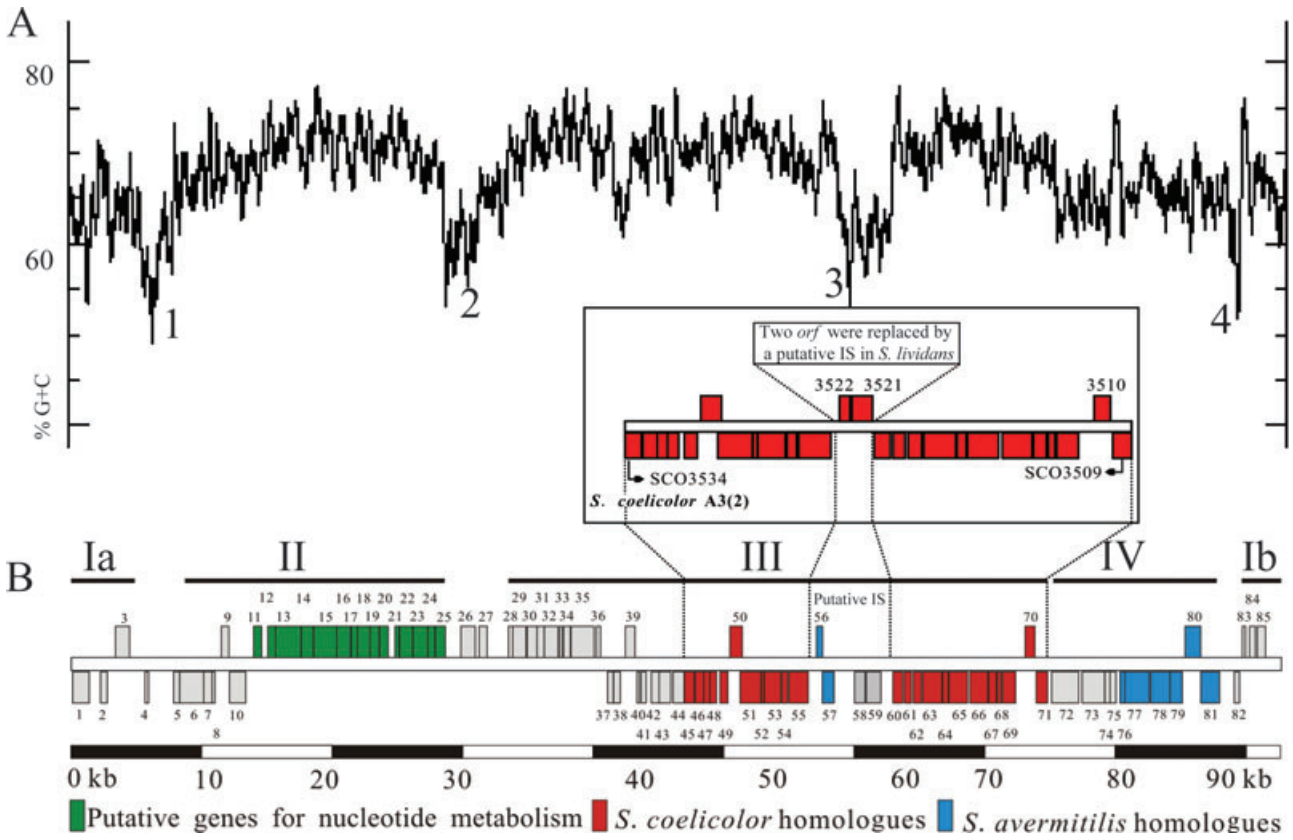


**Fig. 1.** A. G+C plot of the SLG DNA sequence. The numbers next to the troughs in the plot represent the four low G+C content valleys. B. Organizational map of the 85 predicted ORFs on the chromosomally integrated SLG element extending from the left junction *attL* to the right junction *attR*. The four modules (I–IV) referred to in the text are shown as thick lines; boxes above and below the axis represent ORFs in the forward and reverse frames respectively. The ORFs shown in green are predicted to be involved in nucleotide metabolism and biosynthetic pathways; red boxes indicate ORFs with >95% DNA identity to syntenic genes in *S. coelicolor* A3(2); and blue boxes represent homologues of *S. avermitilis*. The inset schematically shows an expanded view of the *S. coelicolor* A3(2) region related to SLG-formation (*slg45–71*) and highlights the limited nature of the modular swap that has occurred. There are no *S. coelicolor* A3(2) SCO3521 and SCO3522 homologues in *S. lividans* 66, while there is another copy of this tandem gene-pair (SCO0874-0875) in *S. coelicolor* A3(2).

### *SLG can spontaneously excise from the* S. lividans *chromosome*

As a putative GI, can SLG excise from the *S. lividans* chromosome? The failure to detect a putatively excised DNA band specific to SLG by PFGE analysis of total DNA of *S. lividans* 66 (Table 2) prompted us to examine this possibility with a more sensitive PCR method using total DNA of *S. lividans* 66 as template, and oligonucleotides flanking both sides of *attL* (Lp1F and Lp1R) and *attR* (Rp2F and Rp1R) as primers (Table S2, Fig. 2). If SLG excises from the chromosome and circularizes between the *attL* and *attR*, primers Rp2F and Lp1R would only amplify a product of 467 bp; likewise, another pair of primer Lp1F and Rp1R targeting to chromosome fusion between *attL* and *attR* would amplify a 725 bp coupled with the SLG excision. Two PCR fragments with the expected sizes were obtained (Fig. 2), demonstrating that the chromosomally integrated SLG was capable of spontaneous excision from the chromosome of the wild-type 66 at low frequency.

To determine the frequency of spontaneous excision of SLG, we randomly selected 1015 colonies of *S. lividans* HXY1 (*dndA::aadA*) (Zhou *et al*., 2005) and patched onto spectinomycin-containing medium after growth on SFM medium to identify putative SLG-minus derivatives. None of the colonies exhibited spectinomycin sensitivity, indicating no SLG was excised. However, when spores were treated with MNNG at a concentration of 1 mg $ml^{-1}$, pH 8 at 30°C for 1 h, as described in Kieser *et al*. (2000), we found that 1 out of 200 (0.5%) randomly selected *S. lividans* HXY1 colonies lost SLG (Fig. S2) via excision at the *attB* site coupled with the loss of resistance to spectinomycin. Therefore, the exposure to MNNG might have increased the frequency of SLG excision in comparison with growth in standard condition. To estimate the rate of SLG excision, we used quantitative real-time PCR analysis and calculated the ratios of the cells containing

 

**Table 1.** *dnd* clusters identified in 11 bacterial strains present in putative GIs.

| Species and strain designation | DNA co-ordinates (predicted ORFs) [contig designation] | Size (kb) | G+C content (%) [genome G+C content] | $\delta^* \times 10^3$ [a] | Genome fragments with lower $\delta^*$ (%)[b] | Other features |
|---|---|---|---|---|---|---|
| *Streptomyces lividans* 66 SLG | 1–92 770 (*slg01–slg85*) | 92.7 | 67.8 [*c.* 70] | 47.5[c] | 100[c] | |
| Ia | 1–8 310 (*slg01–slg05*) | 8.3 | 62.1 | 79.4[c] | 98.4[c] | Integrase (*slg01*) gene, transposase pseudogene (*slg02*) and SLG-associated DR (1–15 bp) |
| II | 8 311–28 683 (*slg06–slg25*) | 20.4 | 65.4 | 80.8[c] | 98.4[c] | Putative nucleotide metabolic pathway |
| III | 28 684–74 926 (*slg26–slg71*) | 46.2 | 68.8 | 41.1[c] | 95.2[c] | Genomic rearrangement, transposase gene (*slg57*) and $\phi$*HAU3*$^r$ |
| IV | 74 927–89 159 (*slg72–slg81*) | 14.2 | 65.1 | 86.7[c] | 99.7[c] | *dnd* cluster |
| Ib | 89 160–92 770 (*slg82–slg85*) | 3.6 | 66.5 | 77.5[c] | 86.4[c] | SLG-associated DR (92 756–92 770 bp) |
| *Streptomyces avermitilis* MA-4680 | 3 655 943–3 684 643 (SCO2922–2937) | 28.7 | 62.8 [70.7] | 76.8 | 99.4 | tRNA-Arg gene, integrase gene (*SAV2736*) and integrase pseudogene (*SAV2737*) |
| *Pseudomonas fluorescens* PfO-1 | 865 302–879 219 (Pfl_0738–0747) | 14.9 | 51.4 [60.5] | 100.9 | 99.1 | tmRNA gene and two integrase genes (*Pfl_0746 , Pfl_0747*) |
| *Pseudoalteromonas haloplanktis* TAC125(ChrII) | 101 629–119 302 (PSHAb0089–0099) | 17.4 | 36.1 [39.4] | 78.1 | 100 | Four conserved genes linked with *dnd* clusters |
| *Escherichia coli* B7A | 9 671–27 566 [NZ_AAJT01000066] 1–33601 [NZ_AAJT01000025] | 54.3 | 49.9 [*c.* 50.8] | 56.6[d] | 100[d] | Insertion sequence elements related to IS*3* and IS66, KPLE2 phage-like element, iron-dicitrate transport gene cluster and three conserved genes linked with *dnd* clusters |
| *Hahella chejuensis* KCTC 2396 | 7 145 272–7 193 152 (HCH_07029–07068) | 47.9 | 47.0 [53.9] | 67.5 | 100 | Tn*7*-like tandem transposase genes |
| *Oceanobacter* sp. RED65 | 38 472–57 669 [NZ_AAQH01000003.1] | 19.2 | 40.7 [*c.* 45] | 76.4[e] | 100[e] | Three conserved genes linked with *dnd* clusters |
| *Bacillus cereus* E33L | 930 024–963 899 (BCZK0812–0835) | 33.9 | 32.6 [35.4] | 60.4 | 99.4 | Three conserved genes linked with *dnd* clusters |
| *Geobacter uraniumreducens* Rf4 | 1–12 996 [NZ_AAON01000085] | 13.6 | 51.4 [54.2] | 65.2[f] | 92.8[f] | Flanking DRs, an island-borne integrase gene and a second identical integrase gene immediately adjacent to the island |
| *Pelagibacter ubique* HTCC1002 | 59 035–97 062 (NZ_AAPV01000002) | 37.9 | 28.0 [*c.* 29.6] | 39.5[g] | 100[g] | *dnd* cluster |
| *Roseobacter denitrificans* Och 114 | 768 003–792 504 (RD1_0795–0808) | 24.5 | 52 [59] | 114.8 | 100 | *dnd* cluster, two clustered transposition protein at right boundary |

**a.** Dinucleotide bias analysis was adapted from the method proposed by Karlin (2001). The value $\delta^*$ denotes the dinucleotide relative abundance difference between the island fragment and the reference genome. The $\delta^*$ value was calculated with the $\delta\rho$–web program (http://deltarho.amc.nl) (van Passel *et al*., 2005). The high $\delta^*$ values of these fragments indicate a likely heterologous origin.
**b.** The percentage distribution of $\delta^*$ is plotted using the $\delta\rho$–web tool with random host genomic fragments of equal length as input sequences (van Passel *et al*., 2005).
**c.** The complete sequence of *S. coelicolor* A3(2) was employed as the reference genome as its genome is closely related to that of *S. lividans* 66 (Kieser *et al*., 1992).
**d.** The complete sequence of *E. coli* K-12 MG1655 was employed as the reference genome as sequencing of the *E. coli* B7A genome is ongoing and 198 contigs are available to date.
**e.** The contig sequence of the ongoing sequenced *Oceanobacter* sp. RED65 genome (NZ_AAQH01000003.1; 215 045 bp) was employed as the reference. The RED65 genome has an estimated size of 3.53 Mb.
**f.** The complete sequence of *Geobacter metallireducens* GS-15 was employed as the reference genome as sequencing of the *Geobacter uraniumreducens* Rf4 genome is ongoing and 189 contigs are available to date.
**g.** The sequence of the largest *Pelagibacter ubique* HTCC1002 contig (NZ_AAPV01000001; 1 012 886 bp) among four available contigs was employed as the reference. The HTCC genome has an estimated size of 1.33 Mb.

excised-SLG verse all cells in the different media including TSBY (34% sucrose), YEME (34% sucrose) and YMG. The copy number of the excised SLG (as determined by PCR using primers P21F and P21R, Table S2) was compared with the copy number of reference locus (3196th nt to 3343rd nt upstream of *attL*, whose copy number was arbitrarily set to 100%) (determined by PCR using primers LC2F and LC2R, Table S2). We observed that the ratios

 

**Table 2.** Bacterial strains and plasmids used in this study.

| Strain or plasmid | Relevant properties[a] | Comment | Source or reference |
|---|---|---|---|
| *S. lividans* | | | |
| 66 | Wild type, Dnd$^+$, ϕHAU3$^r$, SLP2$^+$, SLP3$^+$ | | Feitelson and Hopwood (1983); Zhou *et al*. (2004) |
| JT46 | *rec*-46, *str*-6, *pro*-2, Dnd$^+$, ϕHAU3$^r$ | | Tsai and Chen (1987) |
| ZX1 | JT46 derivative, selected for DNA stability during electrophoresis, *dnd*, ϕHAU3$^s$ | | Zhou *et al*. (2004) |
| HXY1 | 1326 derivatives with insertion of a *aadA* gene into *dndA*, Spc/Str$^R$ | | Zhou *et al*. (2005) |
| HXY7 | ZX1 derivative obtained following conjugal acquisition at 28°C of pJTU1514 | Fig. 3A | This study |
| HXY10 /HXY10-1 | ZX1 derivative containing a chromosomally integrated copy of pJTU1515 | Figs 3C and 4 | This study |
| HXY11 | ZX1 derivative containing a chromosomally integrated copy of pJTU1516 | Fig. 4 | This study |
| HXY12 | ZX1 derivative containing episomoal pJTU1522 and a chromosomally integrated copy of pJTU1520 | Fig. 4 | This study |
| HXY13 | HXY12 derivative obtained following spontaneous loss of pJTU1522 | Fig. 4 | This study |
| HXY16 | *S. lividans* 66 derivative with SLG precisely excised at the *attB* site | Fig. S2 | This study |
| HXY18 | *S. lividans* HXY1-derived mutant, *slg10* is disrupted by *ori* (pUC18)-*tsr* mediated by pJTU1519 | | This study |
| HXY19 | ZX1 derivative with *aac(3)IV* cassette inserted at 17 090th nt upstream of the *attL (*the distal end to *dnd* cluster) of SLG | | This study |
| *S. coelicolor* | | | |
| M145 | Wild type strain | Sequenced | Kieser *et al*. (2000) |
| ZH3 | M145 derivative with SCO3930 partially replaced with *aac(3)IV* | | Li *et al*. (2006) |
| *Escherichia coli* | | | |
| DH5α | F$^-$*rec*A *lac*Z ΔM15 | General cloning host | Hanahan (1983) |
| ET12567 /pUZ8002 | Strain used for conjugation between Cml$^r$, Str$^r$, Tet$^r$, Km$^r$ *E. coli* and *Streptomyces* spp. *rec*F, *dam*, *dcm*, *hsdS*, | | Flett *et al*. (1997) |
| 16C3 | SuperCos1-derived cosmid with insert from *S. lividans* 66 carrying *dnd* and *attR* | Genomic library cosmid | Zhou *et al*. (2004) |
| 16H2 | SuperCos1-derived cosmid with insert from *S. lividans* 66 carrying *attL* | Genomic library cosmid | This study |
| 16C2 | SuperCos1-derived cosmid with insert from *S. lividans* 66 | Genomic library cosmid | This study |
| 16E3 | SuperCos1-derived cosmid with insert from *S. lividans* 66 | Genomic library cosmid | This study |
| pHZ132 | Bifunctional pSG5 derivative, *vph, oriT*, *ori*(pAT153), *bla, tsr, cos, rep*, *ori*(pSG5) | | Bao *et al*. (1997) |
| pSET152 | *aac(3)IV, lacZ, rep$^{pUC}$*, *attÔC31*, *oriT* | | Bierman *et al*. (1992) |
| pOJ260 | *aac(3)IV*, *oriT*, *rep$^{pUC}$*, *lac*Z | | Kieser *et al*. (2000) |
| pMD18-T | pUC18 derivative | | TaKaRa |
| pJTU412 | Shuttle cosmid derived from pHZ1358 (Sun *et al*., 2002), *oriT*, *ori*(ColE1), *bla*, *tsr*, *cos*, *rep*(pIJ101), *ori*(pIJ101) | | Sun *et al*. (unpubl. data) |
| pJTU1510 | pMD18-T containing 722 bp PCR product carrying *attR* from 16C3 | Fig. 3A | This study |
| pJTU1511 | pMD18-T containing the PCR amplicon generated from pJTU1515 using primers P-pJTU1515F and P-pJTU1515R | Fig. 3C | This study |
| pJTU1512 | A 3 575 bp ApaI fragment carrying *attL-int-tnp* from 16H2 was cloned into the unique ApaI site of pSET152 | | This study |
| pJTU1513 | pJTU1512 containing an insert bearing the *oriT* and thermo-sensitive *ori* of pSG5 | | This study |
| pJTU1514 | Mini-island-carrying vector constructed by ligation of a HincII-BamHI fragment carrying *attR* from pJTU1510 with EcoRV-BamHI-cut pJTU1513 | Fig. 3A | This study |
| pJTU1515 | The excised and re-circularized mini-island derived from HXY10 DNA following replicative passage in *E. coli* DH10B | Fig. 3C | This study |
| pJTU1516 | pJTU1515 mini-island derivative lacking *tnp* | Fig. 4 | This study |
| pJTU1517 | The region from 148 nt upstream to 12 nt downstream of the *attP* was removed from pJTU1515 with PstI and NheI digestion, blunted and self-relegated. | Fig. 4 | This study |
| pJTU1518 | SmaI DNA fragment (10 364–14 128) was recovered from 16H2 and ligated with pSET152 that was digested with EcoRV and FspI | | This study |
| pJTU1519 | 6 173 bp *tsr-bla-ori* (pAT153) containing NotI fragment from pHZ132 was inserted into NotI site of pJTU1518 | | This study |
| pJTU1520 | pJTU1515 mini-island derivative lacking *tnp* and *int* | Fig. 4 | This study |
| pJTU1521 | pOJ260 (*Streptomyces* suicide plasmid) containing an intact copy of the *int* gene from pJTU1515 | | This study |
| pJTU1522 | pJTU412 containing an intact *int* gene from pJTU1521 | Fig. 4 | This study |

**a.** *oriT*, origin of transfer of plasmid RK2; *tsr*, thiostrepton-resistance gene; *aac3(IV)*, apramycin-resistance gene; *aadA*, streptomycin/spectinomycin (*str/spc*)-resistance gene; *bla*, ampicillin-resistance gene; *vph*, viomycin-resistance gene; *dnd,* gene cluster encoding DNA degradation phenotype; *int*, integrase of SLG; ϕHAU3$^r$, resistant to phage ϕHAU3.

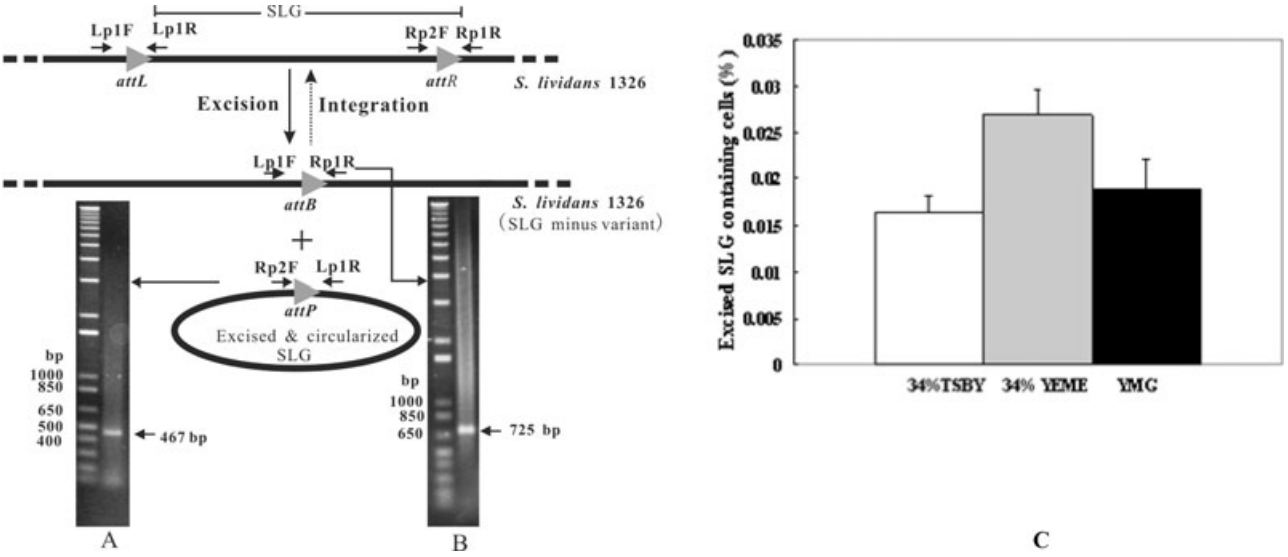


**Fig. 2.** Spontaneous excision of SLG from the chromosome of *S. lividans* 66. Thick straight lines represent the *S. lividans* 66 chromosome; *attL* and *attR* are shown as grey arrowheads, while primer annealing sites are represented by short arrows.
A. PCR analysis of *S. lividans* 66 genomic DNA using primers Lp1F and Rp1R detected the 467 bp fusion fragment between *attL* and *attR*, confirming the existence of excised and circularized forms of SLG.
B. PCR analysis of *S. lividans* 66 genomic DNA using primers Rp2F and Lp1R, which would amplify across the predicted deletion points, detected a 725 bp junction fragment, supporting the existence of an SLG-minus variant in the population.
C. Real-time PCR analysis of the ratios of the cells harbouring the excised SLG. All values were shown as the relative copy numbers to the reference fragment. DNA was prepared from three different media to monitor the difference of excision rates. All data represent the mean ± SD.

of cells harbouring the excised SLG ranged from 0.016% to 0.027% (Fig. 2C) in different media.

### *Role of an integrase in site-specific excision and integration of SLG*

To localize SLG-borne features that might be required for site-specific integration into and/or excision/self-circularization from the chromosome, we constructed a thermo-sensitive *Streptomyces* replicon pSG5-derived (Maas *et al*., 1998) plasmid (pJTU1514) carrying the *attL/attR* sites, an intact integrase gene (*slg01*) and a truncated transposase gene (*slg02*), but with replacement of all other SLG internal genes by pUC18 carrying an *Escherichia coli* replicon and an apramycin-resistance gene (*aac*(*3*)*IV*) (Fig. 3A), using an approach based on the method proposed by Ubeda *et al*. (2003). As a result, we generated a smaller circular molecule with precise site-specific excision via *attL/attR* as a circular SLG-derived mini-island by simple loop-out of pJTU1515 (Fig. 3B) from pJTU1514. (See *Supplementary material*.)

Next, we tested whether pJTU1515 could be site-specifically integrated into the *S. lividans* ZX1 core chromosome ('backbone') at the *attB* site to form a strain identical to HXY10 (Table 2). By protoplast transformation, pJTU1515 was introduced into ZX1, and a strain (HXY10-1) selected by apramycin resistance was subjected to PCR using five primers designed to anneal in the vicinity of the *attL* and *attR* sites: Lp1F and Rp1R annealed to backbone DNA, whereas Lp1R, Rp1F and Rp2F bound to sequences in pJTU1515 (in the directions indicated in Fig. 4). Using primers Lp1F and Lp1R targeting the left junction, a 0.8 kb specific product was amplified from HXY10-1 (Fig. 4A). Similarly, using primers Rp1F and Rp1R that target the right junction, a 0.7 kb specific product was amplified (Fig. 4C). The data suggested that the mini-island had integrated site-specifically at the *attB* site. In addition, the circularly excised mini-island was detected by PCR with primers Rp2F and Lp1R in HXY10-1-derived DNA (Fig. 4B), this suggested that the mini-island existed dynamically in chromosomally integrated and non-replicative, episomal forms in the host strain.

Finally, the three elements at the *murA1*-proximal end of SLG were investigated for their requirement for island excision and integration, including: (i) *slg01 (int)*, encoding an intact integrase homologue, (ii) *slg02 (tnp)*, encoding a likely truncated transposase and (iii) the 15 bp predicted *attP* site. Plasmids derived from pJTU1515 (Fig. 4; Table 2) were constructed and analysed: (i) pJTU1520, in which the 5′ 590 nucleotides of *slg01* and the entire *slg02* were removed from pJTU1515, failed to integrate into ZX1 by transformation, but its integration/excision capability could be restored *in trans* by firstly introducing pJTU1522 carrying an intact *slg01 (int)*, resulting in strain HXY12 (Fig. 4). After a round of non-selective growth of HXY12, the replicative but highly unstable

 

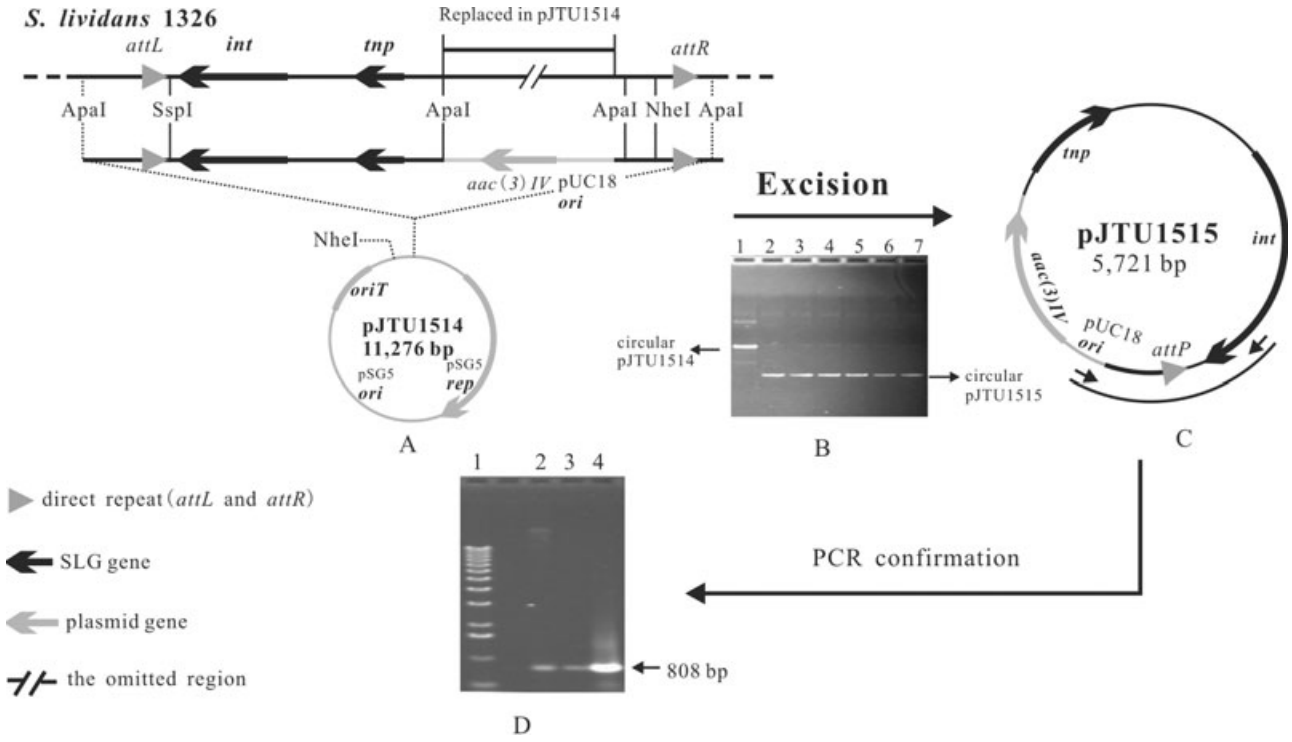


**Fig. 3.** Integration and excision of a SLG mini-island in *S. lividans* 66. The mini-island was introduced on a thermo-sensitive plasmid, pJTU1514; derivatives that had undergone an island-swapping event were selected by growth at 42°C in the presence of apramycin, resistance to which was encoded by the mini-island-borne *aac(3)IV* gene.
A. Schematic representation of the SLG mini-island used in this study and its cognate wild-type island in the chromosome of *S. lividans* 66. The *attL* and *attR* boundaries are shown as grey triangles, *S. lividans* 66 DNA as black lines or arrows, and plasmid DNA or mini-island-specific components (*aac*(*3*)*IV* and pUC18 *ori*) as grey lines or arrows. Dashed lines indicate restriction sites outside the mini-island, while solid lines denote restriction sites within the boundaries of the mini-island.
B. Gel electrophoresis of undigested pJTU1514 (lane 1) and the spontaneously excised and circularized mini-island pJTU1515 (lanes 2–7). The latter plasmid was originally derived from the *S. lividans* ZX1-derivative HXY7 but was passaged through *E. coli* DH10B to allow for pUC18 *ori*-based amplification.
C. Map of pJTU1515 showing the positions of primer binding sites for P-pJTU1515F and P-pJTU1515R (short arrows).
D. PCR analysis using the above primers spanning the recircularization junction of the excised mini-island confirmed the existence of this entity in *S. lividans* HXY10 (lanes 2 and 3); pJTU1515 plasmid DNA was used as a positive control (lane 4).

pHZ1358-derived pJTU1522 was cured from the mycelium, leading to HXY13 (Fig. 4). PCR with primers Lp1R and Rp2F failed to detect free pJTU1520 in HXY13, although all the expected PCR fragments could be detected in HXY11, HXY12 and HXY13 by PCR using the same primers used for HXY10-1 (Fig. 4A–C). (ii) pJTU1516, in which the 3′ 306 nucleotides of *slg02 (tnp)* were deleted from pJTU1515, was introduced by transformation into ZX1 to generate HXY11, the resultant plasmid was found to be capable of integration, excision and circularization, as pJTU1515 (in HXY10-1). (iii) pJTU1517 was constructed by removing a 175 bp *attP* region from pJTU1515, and was found to be still able to integrate into the ZX1 at a similar rate as pJTU1515, the exconjugants were confirmed by PCR using primer pairs: P1F&R, P2F&R and aac(3)IV-T F&R (Table S2, data not shown), suggesting the availability of a secondary *attP* site other than *attP* (15 bp) locating within the pJTU1517. The secondary attachment site is yet to be determined. Hence, the functional integrase encoded by *slg01 (int)* on SLG was found to be necessary to mediate site-specific integration, excision and circularization of the mini-island.

### *The native SLG seems to be non-transmissible in* Streptomyces

To determine whether the SLG island is transmissible between *Streptomyces* strains, an interspecies mating experiment was first performed with *S. lividans* HXY1 (*dndA::aadA*, resistant to streptomycin) harbouring a helper conjugative plasmid pIJ101 (Kieser *et al*., 1982) as a donor, and *S. coelicolor* M145-derived recipient strain ZH3 (*SCO3930::aac*(*3*)*IV*, resistant to apramycin) (Li *et al*., 2006) as a recipient. No ZH3-derived exconjugants with acquired SLG, however, could be obtained, as monitored from an intensive screening of *c.* 100 streptomycin–apramycin double resistant colonies (with production of pigmental actinorhodin characteristic of the recipient strain

 

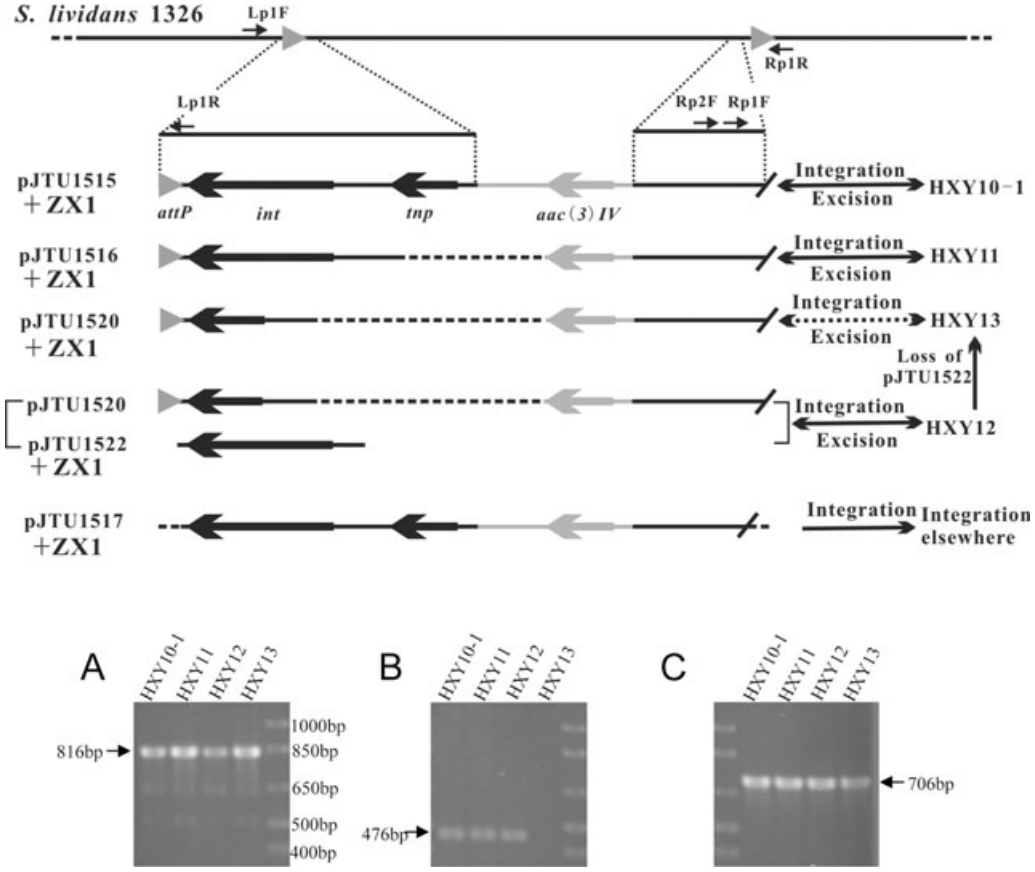


**Fig. 4.** Localization of essential elements for SLG excision and integration by constructing SLG derivatives. The symbols are the same as in Fig. 3. Primer annealing sites are indicated by short arrows. pJTU1515, pJTU1516, pJTU1517 and pJTU1520 are shown in the linear form; dotted lines indicate regions deleted. The replicative plasmid pJTU1522 carried an intact copy of the SLG integrase gene. HXY10-1, HXY11 and HXY13 are ZX1 derivatives containing an integrated copy of pJTU1515, pJTU1516 and pJTU1520 respectively. HXY12, the parent strain of HXY13, contains an integrated copy of pJTU1520 and episomal pJTU1522.
A. PCR analysis using primers Lp1F and Lp1R and genomic DNA as indicated, targeting the left junctions of integrated mini-islands.
B. PCR analysis using primers Rp2F and Lp1R, targeting the fused junctions of excised circular mini-islands. No PCR product was observed with HXY13 DNA.
C. PCR analysis using primers Rp1F and Rp1R, targeting the right junctions of integrated mini-islands.

ZH3) by PCR amplification using a pair of primers (M1F and M1R) targeting to module I of SLG (exemplified in Fig. S3). Meanwhile, in a second intraspecies mating experiment involving use of *S. lividans* HXY18 (*dndA::aadA, slg10::tsr*) with the helper conjugative plasmid pIJ101 as a donor, and HXY19 (ZX1 derivative with *aac*(*3*)*IV* inserted at 17 090th nt upstream of the *attL*) as a recipient, no HXY19-specific exconjugants with acquired SLG was detected. It seemed therefore that the native SLG may have lost its capabilities of inter-, and even intraspecies transmission between *Streptomyces* strains.

### *Widespread distribution of* dnd *clusters among distantly related bacterial species*

Interrogation of bacterial genome sequences (available in GenBank up to October 2006) identified additional likely homologues of the *dnd* cluster in another 11 strains belonging to phylogenetically diverse bacterial species (Fig. 5; Fig. S4). These strains represented species and genera of variable origin and diverse habitats, such as GC-rich *S. lividans* versus AT-rich *Pelagibacter ubique*, soil-dwelling organisms versus marine microbes, non-pathogenic saprophytes versus human pathogens.

Eight of the 12 genomes bearing the *dnd* cluster contained homologues of five genes. Examination of gene order and spacing suggested that *dndA* and *dndB-E* comprised independent transcription units (Fig. 5). Indeed, reverse transcription-PCR on *S. lividans* 66 confirmed that *dndB-E* formed a single operon (J. Liang *et al.* unpubl. data). DndA protein could function as an IscS-like cysteine desulphurase (Schwartz *et al.*, 2000; You *et al.*, 2007) and is essential for Dnd phenotype (Zhou *et al.*, 2005). The *dndA* gene is divergently transcribed relative to *dndB-E* in Strain A, B, D, I and J (Fig. 5) but lies downstream of and in the same orientation as *dndB-E* in

 

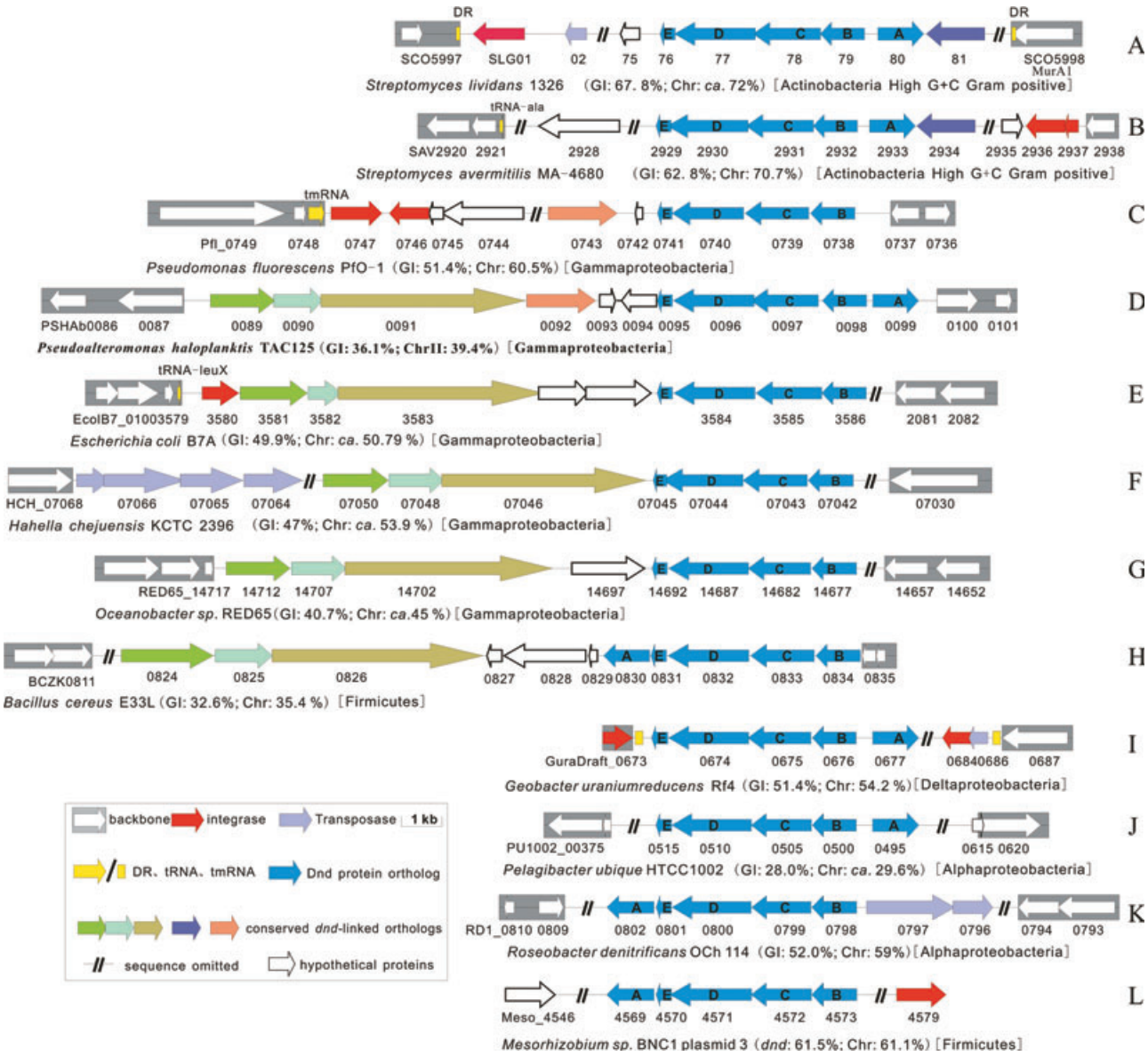


**Fig. 5.** Maps of putative *dnd*-encoding GIs and backbone in 11 bacterial chromosomes and one plasmid. The *dnd* gene clusters are shown as blue arrows, with the labels referring to the matching *dnd* gene in *S. lividans* 66. Red, light purple and yellow arrows indicate genes coding for functionally similar proteins that do not necessarily exhibit sequence similarity, while the remaining filled arrows indicate sets of genes sharing sequence homology. Unfilled arrows indicate genes coding for hypothetical proteins. Hatched boxes represent flanking genomic backbone. The G+C content of islands and chromosomes and the taxonomic division to which the organisms belong are shown in curved and square parentheses respectively. GenBank accession numbers are shown.

H, K and L (Fig. 5). No putative *dndA* orthologue was identified in the immediate vicinity of *dndB-E* in C, E, F and G (Fig. 5). However, an *iscS* homologue was present elsewhere in these bacterial genomes and the cognate proteins may have served as functional homologues of DndA. We are currently exploiting these data to further examine the functional nature of the Dnd proteins.

Notably, five of the 12 strains with a genomic G+C content of less than 56% all contain another highly conserved three-gene cluster tightly linked with the *dnd* cluster (Fig. 5D–H). Additionally, the gene designated PSHAb0092 (Fig. 5D) shares 35% amino-acid identity with Pfl_0743 (Fig. 5C). SAV2928 (Fig. 5B) shares 38% identity with RD1_0805 (Fig. 5K), 28% identity with Meso_4564 (Fig. 5L). Finally, at the right flanking region of the *dnd* loci, Slg81 (Fig. 5A) resembles SAV2934 (Fig. 5B) with 72% amino-acid identity.

To test the prevalence of the *dnd* cluster in the same genera, we performed a Dnd phenotype survey on 74 actinomycete strains collected from geographically distinct

 

regions. The DNA of five *Streptomyces* strains were found to possess Dnd phenotype as that of *S. lividans* 66 and *S. avermitilis* MA-4680 (*S. acrimycini* 2236, *S. caneuscus*, *S. griseoplanes* 92023, *S. verticillus* ATCC 15003 and *Streptomyces* sp. As-01) (data not shown). We also investigated the linkage between *dnd* and integrase in these surveyed actinomycetes strains by Southern hybridization using *slg01* as a probe. Five newly identified Dnd$^+$ strains showed a negative signal, whereas one Dnd$^-$ strain (*Streptomyces* sp. 30214) produced a positive result (data not shown). This observation suggests that *dnd*-bearing GIs do not necessarily have to be linked with a specific integrase gene, but can be equipped with one or more alternatives (Fig. 5). Dnd phenotype has also been discovered recently in other bacterial genus or species including two *E. coli* isolates causing blood stream infections in humans (K. Rajakumar *et al*., unpubl. data). Two isolates of *Salmonella enterica* serovar Livingstone and two isolates of *S. enterica* serovar Cerro have been identified by normal agarose gel electrophoresis to display the Dnd phenotype (Murase *et al*., 2004), whose counterpart gene cluster had been isolated (T. Xu, pers. comm.). Similarly, 11 out of 34 clinical isolates of *Pseudomonas aeruginosa* have the Dnd phenotype (Romling and Tummler, 2000). These findings suggested that *dnd* clusters are more common among different strains than had been anticipated based on available genome sequence data.

### *All* dnd *clusters in other 11 sequenced genomes locate on mobile elements*

As mentioned above, the *S. lividans* 66 *dnd* cluster lies in module IV of the large, mosaic SLG. This discovery prompted us to examine further the genomic context of the *dnd* clusters in the other 11 completely or partially sequenced bacterial strains. Remarkably, all 11 other identified *dnd* clusters seemed to lie on mobile genetic elements, 10 within the chromosome and one on a large plasmid (Fig. 5L, Table 1). Sequence analysis showed that these putative *dnd*-encoding islands shared key features typical of GIs on the basis of G+C content (Column 4, Table 1), dinucleotide bias (Column 5 and 6, Table 1), context homology and GI characteristic elements, such as integration into tRNA, or possession of DRs, integrase and/or transposase (Column 7, Table 1; Fig. 5). The prediction that the *dnd*-bearing fragments are on GIs in *E. coli* B7A and *Hahella chejuensis* KCTC 2396 agrees well with the predication of Ou *et al*. (2007) and Jeong *et al*. (2005) respectively. The fragment encoding the *dnd* homologue in *S. avermitilis* MA-4680 is also a putative island as predicted by ISLANDPATH (Hsiao *et al*., 2003). In addition, *dnd* cluster is also found on the plasmid in *Mesorhizobium* sp. BNC1 (Fig. 5L) or on plasmid-derived chromosome in the *Pseudoalteromonas haloplanktis* TAC125 (Fig. 5D) (Medigue *et al*., 2005). Collectively, we demonstrated that the *dnd* clusters identified in all known cases up to now are located on the mobile elements.

## Discussion

The 93 kb SLG of *S. lividans* 66, absent from its close relative *S. coelicolor* A3(2), was found to integrate into the 3′ terminus of *murA1*, a well-conserved locus in three sequenced *Streptomyces* genomes. Like many functional GIs (Williams, 2002), the direct relevance of *att* and *int* for mediating site-specific excision and integration of SLG into the *S. lividans* 66 chromosome was demonstrated using a constructed mini-island with *int* sandwiched between *attL* and *attR* in a temperature-sensitive plasmid. Across the mosaic-like, four-module SLG, the *dnd* gene cluster is present on module IV, a relatively small portion of this large island. We propose that module I, flanked by *att* sites and including *int,* is the basic mobile entity of SLG. Module II encodes hypothetical proteins involved in a putative nucleotide-related metabolic pathway. Module III shows greater than 95% nucleotide identity to a contiguous region of *S. coelicolor* A3(2); however, two internal open reading frames (ORFs) are replaced by a putative IS element, including a transposase gene and two tightly linked genes, *ea31* and *ea59* (known as $\phi$*HAU3*$^r$, Zhou *et al*., 1994b) in six bacterial species (Fig. S5), such organization is reminiscent of the IS element containing *ea31* and *ea59* in *Pseudomonas syringae pv.* tomato DC3000 (Alfano *et al*., 2000). Such a mosaic compilation of diverse functional modules could involve an initial invasion by a mobile element, containing, for example, a component of module I and followed by recombinational promiscuity in disparate organisms, and suggests that SLG could be an example of how a bacterial host has successfully captured peripatetic genetic information from multiple sources. Thus, given the highly mosaic structure of SLG, we suggest that the present entity represents a relic of multiple past recombination events that may have since lost its native capability of self-transmission.

Horizontal transfer of GIs is often initiated by excision of a linear form from the chromosome to form a circular, mobilizable episome. This process can be induced by a degradable substrate in the growth medium, as in activation of *clc* excision (Sentchilo *et al*., 2003a); regulated in a cell density-dependent manner, as in excision of ICE*Ml*SymR7A (Ramsay *et al*., 2006); or more often by activation of the promoter of an independent excisionase or integrase, which also leads to GI integration into the genome of the recipient strain (Sentchilo *et al*., 2003a; Doublet *et al*., 2005; Ramsay *et al*., 2006). Apparently, the excision of SLG falls into the last case as a result of the absolute necessity of the *int* gene (Fig. 4). The currently reported excision rate of the SLG in different media, which

was kept at a stable level (0.016%~0.027%) after standard 36 h of incubation, is a suggestion that the regulation on the integrase is not so dependent on properties of the medium component, such as osmotic pressure (34% sucrose), and/or the availability of the iron (YMG is a low-iron medium compared with YEME). By contrast, the likely upregulation of GI integrase activity, as detected by the increased SLG excision after MNNG exposure, is reminiscent of increased prophage induction from lysogens when hosts are subjected to UV irradiation (Tomizawa and Ogawa, 1967), but its exact mechanisms have not been investigated any further.

Horizontal gene transfer of the excised GI usually involves process of conjugation, transformation or phage transduction. Conceivably, the conjugation would need genes related to DNA transfer, whose presence was apparently not detected bioinformatically in SLG, in agreement with the detected absence of its self-transmissibility. However, the lack of the self-transmissibility of a GI could also result from the deletion of its transfer functions after conjugation, a case of which had been described in *E. coli* ECOR31 that a 35 kb transfer region of a conjugative plasmid was deleted from the Yps HPI but present on the *E. coli* HPI (Schubert *et al*., 2004). As well, the failed detection of the mobilized transfer of the SLG by the conjugative plasmid pIJ101 does not necessarily conclude that the SLG is non-mobilizable as transfer systems between GI and the helper plasmid have to match well each other, and some of the GIs may need phage transduction system(s) other than conjugal system(s) of the helper plasmid(s) for its/their mobilization. Examples of the mobilization by phage transduction systems had been reported by Ruzin *et al*. (2001) for phage SaPI1, and by O'Shea and Boyd (2002) for phage VPI respectively. In addition, even the mobilization of the GIs by the phage transduction systems has their matching specificities. For example, phages 13 and 80$\alpha$ were reported able to mobilize SaPI1 (Ruzin *et al*., 2001), but failed to mobilize SaPIbov in *Staphylococcus aureus* (Fitzgerald *et al*., 2001). Furthermore, while SaPIbov2 was reported capable of excision from the chromosome *Staphylococcus aureus*, its mobility is still unclear till now (Ubeda *et al*., 2003). Given the above considerations, and the mosaic structure of the SLG, it is tempting to speculate that SLG may have lost genes required for its natural and/or mobilized transfer in order to maintain a relatively stable inheritance with the host during evolution. To our knowledge, no other *Streptomyces* GIs had been described or demonstrated in details of being able to integrate into and/or excise from their chromosomes, as demonstrated here as a result of a specific control by a discrete *int* gene via a localized 15 bp DRs, in SLG.

Although it is yet undefined how the *dnd* gene cluster evolved and was disseminated across different bacterial species based on the limited genome sequence data, the diversity of the *dnd*-bearing hosts, the markedly different *dnd* sequence signatures and the lack of a functionally mobile *dnd* GI, collectively suggest that the *dnd* cluster was organized into a functional locus on a conjugative plasmid and/or other mobile element in very ancient times, prior to extensive spread and sequence diversification over the eons. Plasmids have been proposed to play a role in the evolution and dissemination of some GIs, such as the *exoU* locus of *Pseudomonas aeruginosa* (Kulasekara *et al*., 2006). In this study, the *dnd* cluster islands were found in *Mesorhizobium* sp. BNC1 plasmid 3 (Fig. 5L) and the plasmid-derived Chromosome II of *Pseudoalteromonas haloplanktis* TAC125 (Fig. 5D). In addition, the degree of gene organization conservation extends beyond the *dnd* clusters for several of these GIs (Fig. 5D–H), suggesting these elements might have been evolved from a common ancient ancestor. Similarly, we hypothesize that the plasmid-derived, mosaic-like small chromosome of *P. haloplanktis* TAC125 could have served as a 'natural depository' for various accessory genetic elements, such as the *dnd* loci and its linked genes that could potentially be sourced from diverse bacterial or phage donors (Fig. 5D).

The widespread occurrence of highly conserved *dnd* gene clusters in the bacterial kingdom that are, however, only sporadically represented among members of a species, is reminiscent of classical DNA methylation-based restriction-modification systems, which frequently play a key role in preventing the uptake of foreign DNA or in altering the way in which the genetic blueprint of an organism is decoded and translated into proteins (Brezellec *et al*., 2006). It is reasonable to assume that the sulphur-based DNA modification system in widespread bacterial species conferred by *dnd* clusters on GIs may play a role in DNA maintenance (Bolden *et al*., 1984), replication (Schmidt *et al*., 1985) and/or code for novel DNA uptake barrier systems (Bair and Black, 2006) that ultimately help promote the survival and dispersal of 'selfish islands'. We are currently actively exploring these hypotheses.

## Experimental procedures

### *Bacterial strains and plasmids, growth conditions and genetic manipulations*

The construction of plasmids and strains are listed in Table 2. Culture and standard bacteriological methods were generally as described by Sambrook *et al*. (1989) and Kieser *et al*. (2000). *E. coli* strains were grown at 37°C in Luria–Bertani medium. *Streptomyces* strains were routinely grown at 28°C on SFM (Kieser *et al*., 2000) agar for sporulation or for conjugation between *E. coli* ET12567/pUZ8002 and *Streptomyces*; on R2YE (Kieser *et al*., 2000) for protoplast

 

transformation; on solid MM (Kieser *et al*., 2000) medium for thermo-sensitivity tests; or in YEME and TSBY (Kieser *et al*., 2000) liquid media supplemented with 34% (w/v) sucrose or YMG (glucose 4.0, malt extract 10.0, yeast extract 4.0, pH 7.2) for mycelial growth.

Plasmid and total DNA were prepared from *Streptomyces* strains according to Kieser *et al*. (2000). Unmethylated DNA was prepared from *E. coli* ET12567. *In vivo* generation of targeted mutations in *Streptomyces* was achieved by conjugation between *E. coli* ET12567 containing the RP4 derivative pUZ8002 (Flett *et al*., 1997) and *S. lividans* according to Kieser *et al*. (2000).

### *DNA sequencing and sequence analysis*

Four overlapping cosmids, 16C3, 16C2, 16E3 and 16H2 (Zhou *et al*., 2004), were purified (Sambrook *et al*., 1989) and sheared with a 550 Sonic Dismembrator (Fisher Scientific) into *c.* 2–3 kb fragments. These random fragments were blunted with T4 polymerase (Fermentas) at 11°C for 20 min, gel purified and ligated into pUC18 for shotgun sequencing. Some gaps were filled using PCR amplifications as detailed in Zhang *et al*. (1999). The assembled sequence covers the complete 92 770 bp SLG, flanked by 24 872 bp and 19 223 bp at the left and right boundary respectively.

The nucleotide sequence of SLG and the flanking regions has been deposited in GenBank under accession number EF210454. Putative protein coding sequences larger than 150 bp were predicted using FramePlot beta 3.0 (Ishikawa and Hotta, 1999), with a 120 bp sliding window and a step of 15 bp. Homology searches were performed by using BLAST at National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/blast/). The G+C plot (Fig. 1A) was drawn using the Freak software within the EMBOSS package (Altschul *et al*., 1997) with a 300 bp sliding window and a step of 10 bp. Dinucleotide bias analysis was performed using the method proposed by Karlin (2001). The dinucleotide relative abundance value $\delta^*$ (Karlin, 2001) was calculated with the δρ–web program (http://deltarho.amc.nl/) (van Passel *et al*., 2005).

### *PCR primers and PCR reaction conditions*

Primers used in this study are listed in Table S2. PCR reactions (50 μl) containing 5 ng template DNA, 25 pM of each primer, 1.5 mM $MgCl_2$, 100 μM dNTPs, 1 unit *Taq* polymerase, in 1× PCR buffer (Sangon, Shanghai, China) were performed. PCR cycling conditions were as follows: initial denaturation at 94°C for 180 s, followed by 30 cycles of denaturation at 94°C for 30 s, annealing at a primer-specific temperature for 30 s, and final extension at 72°C for a duration dependent on the length of the expected amplicon (50 s $kb^{-1}$). PCR products were purified from 0.8% agarose gels using the DNA Gel Extraction Kit (V-gene Biotechnology Limited, China) and subsequently inserted into pMD18-T vector (TaKaRa, Dalian, China) for sequencing.

### *Real-time PCR*

Real-time PCR was performed using a Mastercycler® ep realplex 4S (Eppendorf) in the presence of SYBR-green. The PCR was performed using SYBR® Premix ExTaq™ [TakaRa Biotechnology (Dalian) Co. Ltd] according to the manufacturer instruction. Cells of *S. lividans* 66 were harvested after 36 h of standard growth in TSBY (34%), YEME (34%) (Kieser *et al*., 2000) and YMG to investigate whether there exists variation of the excision rate influenced by the media components, i.e. osmotic pressure, ion. Amplification products were designed to be less than 200 bp in size and primers are listed in Table S2. The left flanking fragment (nt-3 343–nt-3 196; LC2F and LC2R) of SLG was used as the reference loci. Primer P21F and P21R (nt-92 686–nt-56) is targeting for quantification of ratios of cells bearing excised SLG; Reactions was performed in 20 μl volumes. PCR condition was set as follows: initial denaturation at 95°C for 5 min, followed by 40 amplification cycles at 95°C for 15 s, and final extension at 60°C for 30 s. Melting-curves were analysed at the end of each elongation step to validate the amplification specificity. All PCR amplifications were performed in duplicates on different days to validate the reproducibility of the assays. The relative copy number for each size of DNA molecule was calculated using the comparative Ct method with the formula $2^{-\Delta\Delta Ct}$ ($\Delta\Delta Ct = \Delta Ct$ sample $- \Delta Ct$ reference). The copy number for reference fragment was assigned a value of 100%, others were presented as calculated percentage relative to the copy numbers of the reference locus.

### *Determination of integrase activity through excision of the mini-island from pJTU1514*

pJTU1514 was introduced into *S. lividans* ZX1, a *dnd*-mutant of *S. lividans* 66 (Zhou *et al*., 2004), by conjugation according to Kieser *et al*. (2000). After cultivation on SFM at 28°C for 16 h, exconjugant HXY7 was transferred to grow at 42°C for 72 h on MM agar supplemented with apramycin (30 μg $ml^{-1}$). These exconjugants were propagated for sporulation on SFM agar at 28°C for 4 days. Total DNA was prepared from mycelium of HXY10 and used to transform *E. coli* DH10B by electroporation. Plasmid DNA was extracted from *E. coli* transformants exhibiting resistance to apramycin and digested with ApaI, NheI, SspI and BglII to determine which part was deleted. Subsequently, the fusion junction was amplified by PCR with primers P-pJTU1515F and P-pJTU1515R and the resulting product inserted into pMD18-T, leading to pJTU1511.

### *Conjugal mating experiments between* Streptomyces

Conjugal mating experiments were carried out using the *S. lividans* HXY1 (*dndA*::*aadA*) or *S. lividans* HXY18 (*dndA*::*aadA*, *slg10*::*tsr*) as the SLG donor strain and the *S. coelicolor* M145 mutant ZH3 (*SCO3930*::*aac*(*3*)*IV*) as a recipient for the interspecies conjugal mating or *S. lividans* HXY19 (ZX1 derivative with *aac*(*3*)*IV* inserted at 17 090th nt upstream of the *attL*) as a recipient for the intraspecies conjugal mating. Spores of both strains were mixed and cocultured on the SFM medium at 28°C for 20 h. The mating plates were overlaid with 50 μg $ml^{-1}$ streptomycin, 100 μg $ml^{-1}$ apramycin and/or 15 μg $ml^{-1}$ thiostrepton to select the potential exconjugants. The donor-to-recipient ratios range from 10:1, 1:1 and 1:10. The number of either donor or recipient spores was approximately $10^7$–$10^9$ cfu per 90 mm plate on SFM agar.

 

**Acknowledgements**

We thank Prof Sir. D. A. Hopwood, FRS for critical reading of the manuscript and many valuable comments. We are grateful to Dr James M. Fleckenstein of University of Tennessee Health Sciences Center, USA, for providing *E. coli* B7A and the Institute for Genomic Research (TIGR) for their policy of making preliminary sequence data publicly available and acknowledge the use in this study of unpublished genome data corresponding to *E. coli* B7A. This work received support from the National Science Foundation of China, the 863 and 973 programs from the Ministry of Science and Technology, the Funds from the Ministry of Education and the Shanghai Municipal Council of Science and Technology.

**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 USA* **97:** 4856–4861.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

Bair, C.L., and Black, L.W. (2006) A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. *J Mol Biol* **366:** 768–788.

Bao, K., Hu, Z., Zhou, X., Zhou, Q., and Deng, Z. (1997) A bifunctional cosmid vector for the mobilized conjugal transfer of DNA from *E. coli* to *Streptomyces* sp. *Prog Nat Sci* **7:** 568–572.

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., *et al.* (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417:** 141–147.

Bierman, M., Logan, R., O'Brien, K., Seno, E.T., Rao, R.N., and Schoner, B.E. (1992) Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* **116:** 43–49.

Bolden, A., Ward, C., Siedlecki, J.A., and Weissbach, A. (1984) DNA methylation. Inhibition of *de novo* and maintenance methylation *in vitro* by RNA and synthetic polynucleotides. *J Biol Chem* **259:** 12437–12443.

Brezellec, P., Hoebeke, M., Hiet, M.S., Pasek, S., and Ferat, J.L. (2006) DomainSieve: a protein domain-based screen that led to the identification of dam-associated genes with potential link to DNA maintenance. *Bioinformatics* **22:** 1935–1941.

Choulet, F., Aigle, B., Gallois, A., Mangenot, S., Gerbaud, C., Truong, C., *et al.* (2006) Evolution of the terminal regions of the *Streptomyces* linear chromosome. *Mol Biol Evol* **23:** 2361–2369.

Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2:** 414–424.

Doublet, B., Boyd, D., Mulvey, M.R., and Cloeckaert, A. (2005) The *Salmonella* genomic island 1 is an integrative mobilizable element. *Mol Microbiol* **55:** 1911–1924.

Dyson, P., and Evans, M. (1998) Novel post-replicative DNA modification in *Streptomyces*: analysis of the preferred modification site of plasmid pIJ101. *Nucleic Acids Res* **26:** 1248–1253.

Feitelson, J.S., and Hopwood, D.A. (1983) Cloning of a *Streptomyces* gene for an O-methyltransferase involved in antibiotic biosynthesis. *Mol Gen Genet* **190:** 394–398.

Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., *et al.* (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* **34:** D247–D251.

Fitzgerald, J.R., Monday, S.R., Foster, T.J., Bohach, G.A., Hartigan, P.J., Meaney, W.J., and Smyth, C.J. (2001) Characterization of a putative pathogenicity island from bovine *Staphylococcus aureus* encoding multiple superantigens. *J Bacteriol* **183:** 63–70.

Flett, F., Mersinias, V., and Smith, C.P. (1997) High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting *Streptomycetes*. *FEMS Microbiol Lett* **155:** 223–229.

Gunes, G., Smith, B., and Dyson, P. (1999) Genetic instability associated with insertion of IS6100 into one end of the *Streptomyces lividans* chromosome. *Microbiology* **145 (**Part 9**):** 2203–2208.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166:** 557–580.

Hochhut, B., and Waldor, M.K. (1999) Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfc*. *Mol Microbiol* **32:** 99–110.

Hochhut, B., Wilde, C., Balling, G., Middendorf, B., Dobrindt, U., Brzuszkiewicz, E., *et al.* (2006) Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic *Escherichia coli* strain 536. *Mol Microbiol* **61:** 584–595.

Hopwood, D.A. (2007) *Streptomyces in Nature and Medicine: The Antibiotic Makers.* New York: Oxford University Press.

Hsiao, W., Wan, I., Jones, S.J., and Brinkman, F.S. (2003) IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* **19:** 418–420.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., *et al.* (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21:** 526–531.

Ishikawa, J., and Hotta, K. (1999) 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.

Jeong, H., Yim, J.H., Lee, C., Choi, S.H., Park, Y.K., Yoon, S.H., *et al.* (2005) Genomic blueprint of *Hahella chejuensis*, a marine microbe producing an algicidal agent. *Nucleic Acids Res* **33:** 7066–7073.

Karlin, S. (2001) Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes. *Trends Microbiol* **9:** 335–343.

Kers, J.A., Cameron, K.D., Joshi, M.V., Bukhalid, R.A., Morello, J.E., Wach, M.J., *et al.* (2005) A large, mobile pathogenicity island confers plant pathogenicity on *Streptomyces* species. *Mol Microbiol* **55:** 1025–1033.

 

Kieser, T., Hopwood, D.A., Wright, H.M., and Thompson, C.J. (1982) pIJ101, a multi-copy broad host-range *Streptomyces* plasmid: functional analysis and development of DNA cloning vectors. *Mol Gen Genet* **185:** 223–228.

Kieser, T., Bibb, M.J., Chater, K.F., Butter, M.J., and Hopwood, D.A. (2000) *Practical Streptomyces Genetics. A Laboratory Manual*. Norwich: John Innes Foundation.

Kieser, H.M., Kieser, T., and Hopwood, D.A. (1992) A combined genetic and physical map of the *Streptomyces coelicolor* A3(2) chromosome. *J Bacteriol* **174:** 5496–5507.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305:** 567–580.

Kulasekara, B.R., Kulasekara, H.D., Wolfgang, M.C., Stevens, L., Frank, D.W., and Lory, S. (2006) Acquisition and evolution of the *exoU* locus in *Pseudomonas aeruginosa*. *J Bacteriol* **188:** 4037–4050.

Leblond, P., Redenbach, M., and Cullum, J. (1993) Physical map of the *Streptomyces lividans* 66 genome and comparison with that of the related strain *Streptomyces coelicolor* A3(2). *J Bacteriol* **175:** 3422–3429.

Li, W., Wu, J., Tao, W., Zhao, C., Wang, Y., He, X., *et al.* (2006) A genetic and bioinformatic analysis of *Streptomyces coelicolor* genes containing TTA codons, possible targets for regulation by a developmentally significant tRNA. *FEMS Microbiol Lett* **266:** 20–28.

Liang, J., Wang, Z., He, X., Li, J., Zhou, X., and Deng, Z. (2007) DNA modification by sulfur: analysis of the sequence recognition specificity surrounding the modification sites. *Nucleic Acids Res* **35:** 2944–2954.

Maas, R.M., Gotz, J., Wohlleben, W., and Muth, G. (1998) The conjugative plasmid pSG5 from *Streptomyces ghanaensis* DSM 2932 differs in its transfer functions from other *Streptomyces* rolling-circle-type plasmids. *Microbiology* **144** (Part 10)**:** 2809–2817.

Mantri, Y., and Williams, K.P. (2004) Islander: a database of integrative islands in prokaryotic genomes, the associated integrases and their DNA site specificities. *Nucleic Acids Res* **32:** D55–D58.

Medigue, C., Krin, E., Pascal, G., Barbe, V., Bernsel, A., Bertin, P.N., *et al.* (2005) Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res* **15:** 1325–1335.

Murase, T., Nagato, M., Shirota, K., Katoh, H., and Otsuki, K. (2004) Pulsed-field gel electrophoresis-based subtyping of DNA degradation-sensitive *Salmonella enterica* subsp. enterica serovar Livingstone and serovar Cerro isolates obtained from a chicken layer farm. *Vet Microbiol* **99:** 139–143.

O'Shea, Y.A., and Boyd, E.F. (2002) Mobilization of the *Vibrio* pathogenicity island between *Vibrio cholerae* isolates mediated by CP-T1 generalized transduction. *FEMS Microbiol Lett* **214:** 153–157.

Ou, H.Y., Smith, R., Lucchini, S., Hinton, J., Chaudhuri, R.R., Pallen, M., *et al.* (2005) ArrayOme: a program for estimating the sizes of microarray-visualized bacterial genomes. *Nucleic Acids Res* **33:** e3.

Ou, H.Y., Chen, L.L., Lonnen, J., Chaudhuri, R.R., Thani, A.B., Smith, R., *et al.* (2006) A novel strategy for the identification of genomic islands by comparative analysis of the contents and contexts of tRNA sites in closely related bacteria. *Nucleic Acids Res* **34:** e3.

Ou, H.Y., He, X., Harrison, E.M., Kulasekara, B.R., Thani, A.B., Kadioglu, A., *et al.* (2007) MobilomeFINDER: web-based tools for in silico and experimental discovery of bacterial genomic islands. *Nucleic Acids Res* (in press). doi: 10.1093/nar/gkm/380.

van Passel, M.W., Luyf, A.C., van Kampen, A.H., Bart, A., and van der Ende, A. (2005) Deltarho-web, an online tool to assess composition similarity of individual nucleic acid sequences. *Bioinformatics* **21:** 3053–3055.

Piel, J., Hofer, I., and Hui, D. (2004) Evidence for a symbiosis island involved in horizontal acquisition of pederin biosynthetic capabilities by the bacterial symbiont of *Paederus* fuscipes beetles. *J Bacteriol* **186:** 1280–1286.

Ramsay, J.P., Sullivan, J.T., Stuart, G.S., Lamont, I.L., and Ronson, C.W. (2006) Excision and transfer of the *Mesorhizobium loti* R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS. *Mol Microbiol* **62:** 723–734.

Romling, U., and Tummler, B. (2000) Achieving 100% typeability of *Pseudomonas aeruginosa* by pulsed-field gel electrophoresis. *J Clin Microbiol* **38:** 464–465.

Ruzin, A., Lindsay, J., and Novick, R.P. (2001) Molecular genetics of SaPI1 – a mobile pathogenicity island in *Staphylococcus aureus*. *Mol Microbiol* **41:** 365–377.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor.

Schaffer, A.A., Aravind, L., Madden, T.L., Shavirin, S., Spouge, J.L., Wolf, Y.I., *et al.* (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res* **29:** 2994–3005.

Schmidt, M., Wolf, S.F., and Migeon, B.R. (1985) Evidence for a relationship between DNA methylation and DNA replication from studies of the 5-azacytidine-reactivated allocyclic X chromosome. *Exp Cell Res* **158:** 301–310.

Schubert, S., Dufke, S., Sorsa, J., and Heesemann, J. (2004) A novel integrative and conjugative element (ICE) of *Escherichia coli*: the putative progenitor of the *Yersinia* high-pathogenicity island. *Mol Microbiol* **51:** 837–848.

Schwartz, C.J., Djaman, O., Imlay, J.A., and Kiley, P.J. (2000) The cysteine desulfurase, IscS, has a major role in *in vivo* Fe-S cluster formation in *Escherichia coli*. *Proc Natl Acad Sci USA* **97:** 9009–9014.

Sentchilo, V., Ravatn, R., Werlen, C., Zehnder, A.J., and van der Meer, J.R. (2003a) Unusual integrase gene expression on the *clc* genomic island in *Pseudomonas* sp. strain B13. *J Bacteriol* **185:** 4530–4538.

Sentchilo, V., Zehnder, A.J., and van der Meer, J.R. (2003b) Characterization of two alternative promoters for integrase expression in the *clc* genomic island of *Pseudomonas* sp. strain B13. *Mol Microbiol* **49:** 93–104.

Sullivan, J.T., and Ronson, C.W. (1998) Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc Natl Acad Sci USA* **95:** 5145–5149.

Sun, Y., Zhou, X., Liu, J., Bao, K., Zhang, G., Tu, G., *et al.* (2002) '*Streptomyces nanchangensis*', a producer of the

 

insecticidal polyether antibiotic nanchangmycin and the antiparasitic macrolide meilingmycin, contains multiple polyketide gene clusters. *Microbiology* **148:** 361–371.

Tomizawa, J., and Ogawa, T. (1967) Effect of ultraviolet irradiation on bacteriophage lambda immunity. *J Mol Biol* **23:** 247–263.

Tsai, J.F., and Chen, C.W. (1987) Isolation and characterization of *Streptomyces lividans* mutants deficient in intraplasmid recombination. *Mol Gen Genet* **208:** 211–218.

Ubeda, C., Tormo, M.A., Cucarella, C., Trotonda, P., Foster, T.J., Lasa, I., and Penades, J.R. (2003) Sip, an integrase protein with excision, circularization and integration activities, defines a new family of mobile *Staphylococcus aureus* pathogenicity islands. *Mol Microbiol* **49:** 193–210.

Volff, J.N., and Altenbuchner, J. (1998) Genetic instability of the *Streptomyces* chromosome. *Mol Microbiol* **27:** 239–246.

Williams, K.P. (2002) Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res* **30:** 866–875.

You, D., Wang, L., Yao, F., Zhou, X., and Deng, Z. (2007) A novel DNA modification by sulfur: DndA is a NifS-like cysteine desulfurase capable of assembling DndC as an iron-sulfur cluster protein in *Streptomyces lividans*. *Biochemistry* **46:** 6126–6133.

Zhang, J., Voss, K.O., Shaw, D.F., Roos, K.P., Lewis, D.F., Yan, J., *et al.* (1999) A multiple-capillary electrophoresis system for small-scale DNA sequencing and analysis. *Nucleic Acids Res* **27:** e36.

Zhou, X., Deng, Z., Firmin, J.L., Hopwood, D.A., and Kieser, T. (1988) Site-specific degradation of S*treptomyces 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. (1994a) *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., Deng, Z., Hopwood, D.A., and Kieser, T. (1994b) Characterization of ϕHAU3, a broad-host-range temperate *Streptomyces* phage, and development of phasmids. *J Bacteriol* **176:** 2096–2099.

Zhou, X., He, X., Li, A., Lei, F., Kieser, T., and Deng, Z. (2004) *Streptomyces coelicolor* A3(2) lacks a genomic island present in the chromosome of *Streptomyces lividans* 66. *Appl Environ Microbiol* **70:** 7110–7118.

Zhou, X., He, X., Liang, J., Li, A., Xu, T., Kieser, T., *et al.* (2005) A novel DNA modification by sulphur. *Mol Microbiol* **57:** 1428–1438.

## Supplementary material

The following supplementary material is available for this article:

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

**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.

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

**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.

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

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

**Table S2.** Oligonucleotide primers used in this study.

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2958.2007.05846.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.