The Journal of Microbiology (2011) Vol. 49, No. 4, pp. 684-688
DOI 10.1007/s12275-011-1262-3

NOTE

Complete Sequence and Organization of the *Sphingobium chungbukense* DJ77 pSY2 Plasmid

Sun-Mi Yeon and Young-Chang Kim*

Department of Microbiology, Chungbuk National University, Chungbuk 361-763, Republic of Korea

(Received May 23, 2011 / Accepted June 30, 2011)

***Sphingobium chungbukense* DJ77 is capable of metabolizing priority chemicals of human health concern such as polycyclic aromatic hydrocarbons (PAHs), extracellular polysaccharide (EPS), and antibiotics. Here, we report the complete DNA and genetic organization of the plasmid pSY2 from strain DJ77. A DNA sequence analysis revealed that pSY2 comprises 18,779 bp encoding 22 open reading frames (ORFs) with 59.5% G+C content. The ORFs on pSY2 were classified into DNA replication, conjugative function, transposition, plasmid stability/partition, and other functional groups (transport, fatty acid biosynthesis, stress, and growth rate regulation). Three ORFs on pSY2 were hypothetical proteins.**

***Keywords*:** plasmid, pSY2, characterization, *Sphingobium chungbukense* DJ77

The fate of bacterial plasmids in both nature and the laboratory is of great concern because of their genetic diversity, which confers a variety of ancillary physiological functions that can be shared among populations by various mechanisms. Typical examples of plasmid-borne functions are degradation of recalcitrant chemicals, antibiotic production and resistance, and virulence factors, which have an impact on human health (Tisa and Rosen, 1990). Additionally, plasmids have been evaluated as a promising molecular approach for use in synthetic biology, medicine, ecology, and evolution as well as basic research in molecular and structural biology (Voss, 2007; Azuma *et al*., 2009).

Sphingobium chungbukense DJ77 was originally isolated from chemically contaminated freshwater sediment based on its ability to degrade polycyclic aromatic hydrocarbons (PAHs) and its resistance to streptomycin (Shin *et al*., 1997; Pal *et al*., 2005). Strain DJ77 has been further studied for its ability to degrade or transform recalcitrant toxic compounds, such as toluene, benzoate, biphenyl, anthracene, and phenanthrene (Kim *et al*., 2000; Basta *et al*., 2004). In addition, strain DJ77 has been a focus due to its capacity to produce large quantities of exopolysaccharides (EPS) and glycosphingolipids for food or pharmaceutical use (Burenjargal *et al*., 2007). A genome sequencing project for *S. chungbukense* DJ77 (Chungbuk National University, Korea), which has two chromosomes and six plasmids, is currently in progress. While plasmid vectors are very useful tools for such genetic analyses, there is, until now, no known vector plasmid for *Sphingomonadales*, except for pSY3 of *S. chungbukense* DJ77 (Yeon and Kim, 2009) and pUT1, pUT2, and pCHQ1 of *S. japonicum* UT26S (Nagata *et al*., 2010). In this study, we present the sequence and an analysis of the pSY2 plasmid and provide insight into the plasmid-borne contribution of the host, *S. chungbukense* DJ77. The 'Materials and Methods' used in this study, including bacterial strains and cultures, preparation of genomic DNA, and sequence determination and analysis, can be found in Yeon *et al*. (2008).

We identified the origin (*oriV*) and terminus (*terV*) of vegetative replication using the GenSkew program (http://mips.gsf.de/services/analysis/genskew). We obtained a complete nucleotide sequence of pSY2 generated using ABI 3700 and Genome Sequencer FLX automatic sequencers (Koumi *et al*., 2004; Goldberg *et al*., 2006), and the gaps between contigs were closed by PCR products, which were amplified from genomic DNA using designed primers.

The complete DNA sequence of plasmid pSY2 was obtained using a whole-plasmid shotgun method, and the result was a single contiguous sequence of 18,779 bp. The genome sequences have been deposited in GenBank under accession number JN180627. The average G+C content of pSY2 was 59.5%, a value similar to that of other sphingomonad genomes (60.7-67.8%). The *oriV* was located (91 bp) on the probable mobilization protein C, whereas *terV* was located (16,039 bp) on the transposase Tn3 family protein. A map of pSY2 has been created with the data obtained from the analysis of the ORFs and G+C content by GenomeViz retrieve COG functional categories to enable rapid visualization and subsequent comparisons of several microbial genomes (Ghai *et al*., 2004). The visualization result was compared with a pSY3 map (Fig. 1). Analysis of the circular plasmid revealed that pSY2 contained 22 ORFs, which were classified into six groups according to their predicted functions (Tables 1 and 2). Of the 22 potential ORFs, six (27%) are related to transposition regulation (ORF9.2, 18.2, 19.2, 20.2, 21.2, and 22.2) (Rurherford *et al*., 2007), three ORFs (14%) were highly ho-

* For correspondence. E-mail: youngkim@chungbuk.ac.kr; Tel.: +82-43-261-2302; Fax: +82-43-264-9600

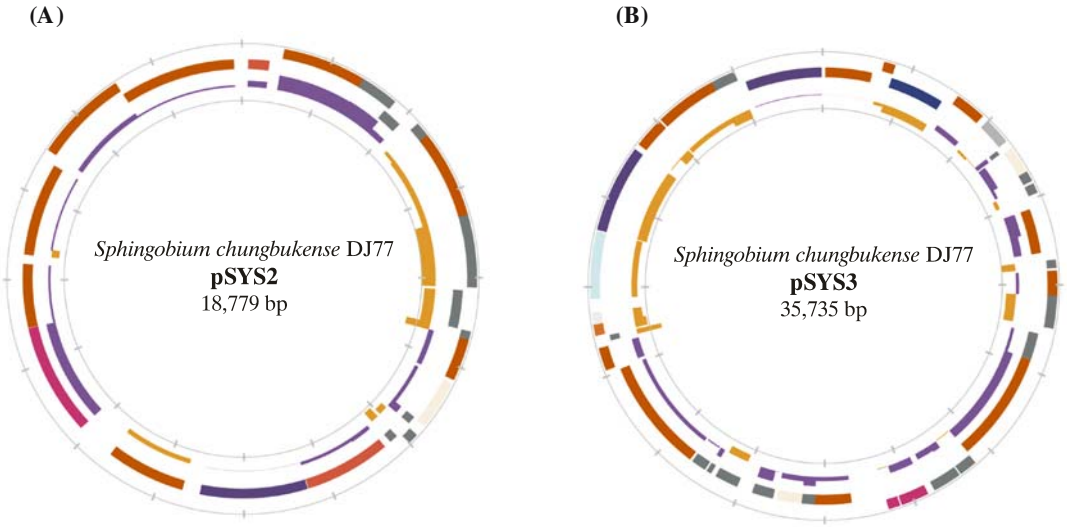

Fig. 1. A map of pSY2 (A) and pSY3 (B) generated by GenomeViz to retrieve COG functional category. The first and second represent the open reading frames (ORFs) of the predicted strain DJ77 on the + and - strands, respectively, colored by role categories: dark orange 1, transcription [K]; dark orange 3, DNA replication, recombination, and repair [L]; antique white, cell division, and chromosome partitioning [D]; tomato, signal transduction mechanisms [T]; deep pink, intracellular trafficking, secretion, and vesicular transport [U]; blue 1, carbohydrate transport and metabolism [G]; light blue 1, coenzyme metabolism [H]; medium purple 4, inorganic ion transport and metabolism [P]; gray 90, general function prediction only [R]; gray 70 [S], not in any COG. The innermost circles represent the GC contents of *S. chungbukense* DJ77 (one circle, yellow, and purple).

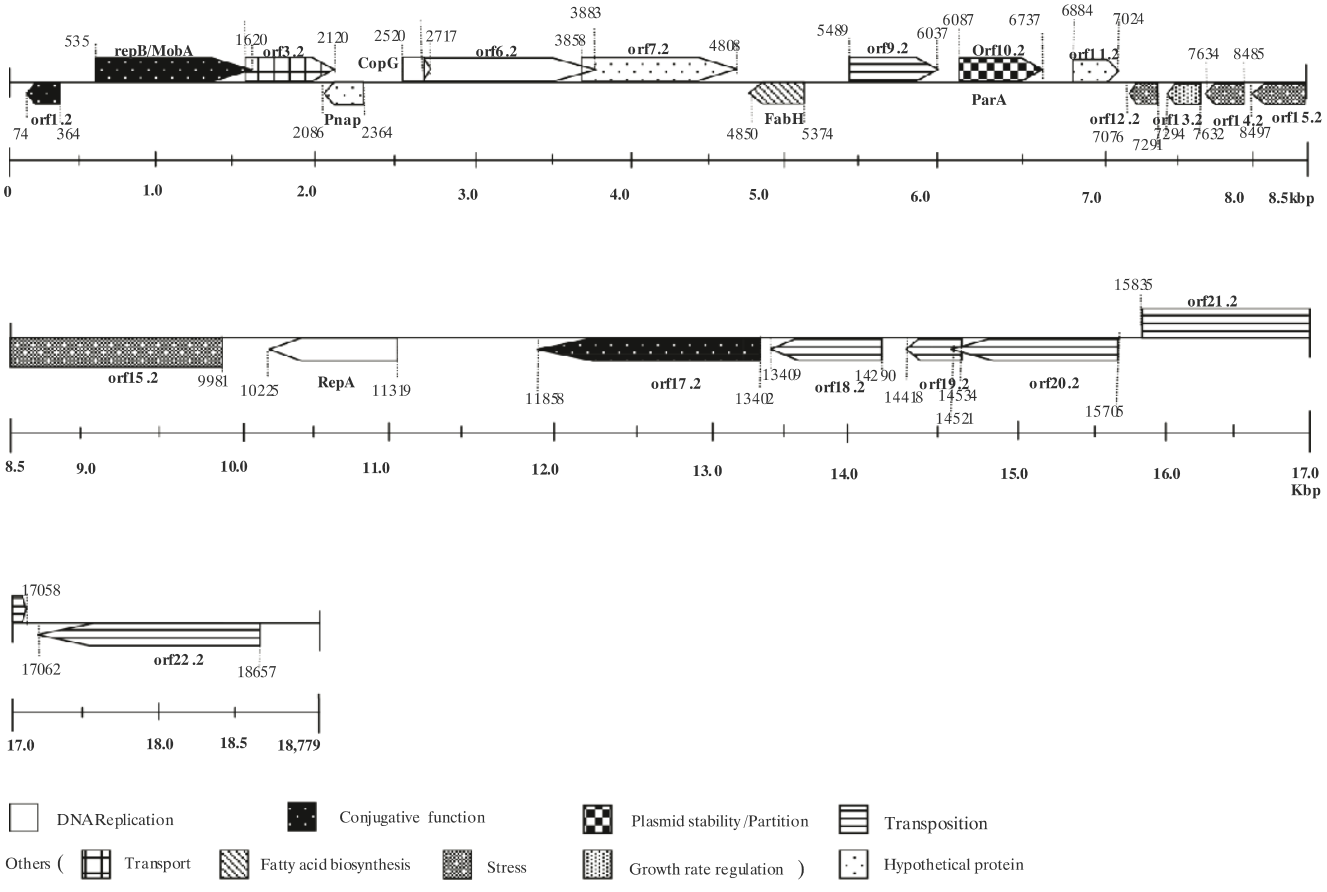

Fig. 2. Genetic organization of the pSY2 plasmid. Numbers shown are the coding sequences from the initiation codon to the stop codon.

Table 1. Putative open reading frames (ORFs) on pSY2 and their functions

ORF	Gene position (nucleotide)		Strand	Length putative protein (aa)	Putative function, closest match	GenBank accession no. of closest match	Alignment region of pSY2 ORF (aa)	Alignment region of closest match (size of homologue) (aa)	% Identity
	Start	End							
1.2	364	74	-	97	Probable mobilization protein C [*Agrobacterium tumefaciens*]	YP_086776.1	118-267	38-89 (107)	46
2.2	535	1620	+	362	repB/MobA-like protein [*Xylella fastidiosa*]	AAK13432.1	1-1029	1-330 (714)	42
3.2	1617	2120	+	168	PREDICTED: solute carrier organic anion transporter family, member 5A1 [*Taeniopygia guttata*]	XP_002197708.1	388-224	2-51 (796)	47
4.2	2364	2086	-	147	Hypothetical protein SJA_P2-00380 [*Sphingobium japonicum* UT26S]	YP_003543440.1	16-276	5-84 (94)	51
5.2	2520	2717	+	66	CopG domain protein DNA-binding domain protein [*Acidiphilium cryptum* JF-5]	YP_001233929.1	1-195	26-90 (90)	44
6.2	2714	3883	+	390	DNA methylase, C-5 cytosine-specific family [*Ruegeria pomeroyi* DSS-3	YP_166300.1	22-1095	4-362 (373)	64
7.2	3858	4808	+	312	Conserved hypothetical protein [*Xylella fastidiosa* Dixon]	ZP_00650809.1	40-627	2-200 (283)	28
8.2	5374	4850	-	175	Beta-ketoacyl-acyl carrier protein synthase III (FabH) [delta proteobacterium MLMS-1]	ZP_01288291.1	489-352	115-155 (346)	44
9.2	5489	6037	+	183	Resolvase [*Rhodopseudomonas palustris* CGA009]	NP_950206.1	1-537	1-179 (183)	61
10.2	6087	6737	+	216	ParA-like protein [*Sphingobium japonicum* UT26S]	YP_003547008.1	1-648	1-217 (217)	94
11.2	6884	7024	+	46	Hypothetical protein SJA_P1-00340 [*Sphingobium japonicum* UT26S]	YP_003547009.1	1-138	59-104 (104)	100
12.2	7291	7076	-	72	Cold-shock DNA-binding domain protein [*Sphingopyxis alaskensis* RB2256]	YP_617386.1	19-204	182-244 (249)	55
13.2	7632	7294	-	113	Transcriptional regulators, TraR/DksA family [*Sphingopyxis alaskensis* RB2256]	YP_617824.1	64-327	22-109 (112)	69
14.2	8485	7634	-	282	Universal stress protein UspA [*Methylobacterium extorquens* DM4]	YP_003068177.1	5-853	1-282 (282)	73
15.2	9981	8497	-	495	Sulphate transporter [*Sphingopyxis alaskensis* RB2256]	YP_617822.1	1-1482	1-494 (494)	75
16.2	11319	10225	-	365	RepA [*Gluconobacter oxydans* 621H, *Sphingobium japonicum* UT26S]	YP_190424.1	1-1092	1-366 (366)	86
17.2	13402	11858	-	515	Probable TraG conjugal transfer transmembrane protein [*Agrobacterium tumefaciens*]	YP_086775.1	196-1536	110-552 (554)	64
18.2	14290	13409	-	289	Resolvase, N-terminal domain [*Sphingomonas wittichii* RW1]	YP_001260446.1	1-879	21-313 (313)	99
19.2	14534	14418	-	39	Transposase Tn3 family protein [*Sphingomonas wittichii* RW1]	YP_001260447.1	1-114	932-969 (969)	100
20.2	15705	14521	-	395	Transposase Tn3 family protein [*Sphingomonas wittichii* RW1]	YP_001260447.1	1-1125	541-915 (969)	100
21.2	15835	17058	+	408	Transposase, mutator type [*Sphingomonas wittichii* RW1]	YP_001260485.1	1-1221	1-407 (407)	99
22.2	18657	17062	-	532	Transposase Tn3 family protein [*Sphingomonas wittichii* RW1]	YP_001260447.1	1-1563	1-521 (969)	97

mologous to proteins for mobilization function and conjugative transfer (ORF1.2, 2.2, and 17.2), and three ORFs (14%) were homologous to known DNA replication-associated proteins (ORF5.2, 6.2, and 16.2), respectively. One ORF (4%) was found for plasmid stability/partitioning (ORF10.2), and six ORFs (27%) had other related cellular functions such as transport, fatty acid biosynthesis, stress, growth rate regulation (ORF3.2, 8.2, 12.2, 13.2, 14.2, and 15.2). Three ORFs (14%) had similarity to hypothetical proteins (ORF4.2, 7.2, and 11.2). As shown in Fig. 2, ORFs on pSY2 were asymmetrically distributed; nine ORFs had a clockwise orientation, and 13 had a counterclockwise orientation. Similar to many other plasmids, pSY2 appeared to be a mosaic consisting of genes from multiple sources.

A comparison using the BLAST program revealed that the putative product of the pSY2-encoded ORF10.2 was a 216-aa protein that had a 94% identity with the ParA-like protein, which mediates site-specific recombination to resolve plasmid

Table 2. Categorization of open reading frames (ORFs) found in pSY2

Sl. No.	Category	ORFs
1	DNA Replication	ORF5.2 (CopG like protein), ORF6.2 (DNA methylase), ORF16.2 (RepA)
2	Conjugative function	ORF1.2 (mobilization protein C), ORF2.2 (RepB/MobA), ORF17.2 (TraG conjugal transfer transmembrane protein)
3	Transposition	ORF9.2 (resolvase), ORF18.2 (resolvase), ORF19.2 (transposase Tn3 family protein), ORF20.2 transposase Tn3 family protein), ORF21.2 (transposase), ORF22.2 (transposase Tn3 family protein)
4	Plasmid stability/Partition	ORF10.2 (ParA-like protein)
5	Others (transport, fatty acid biosynthesis, stress, growth rate regulation)	ORF3.2 (solute carrier organic anion transporter family), ORF8.2 (Beta-ketoacyl-acyl carrier protein synthase III, FabH), ORF13.2 (transcriptional regulators, TraR/DksA family), ORF12.2 (cold-shock DNA-binding domain protein), ORF14.2 (Universal stress protein UspA), ORF15.2 (sulphate transporter),
6	Hypothetical protein	ORF4.2 (Pnap_1147), ORF7.2 (conserved hypothetical protein), ORF11.2 (GDI_3856)

multimers of *S. japonicum*. Therefore, pSY2 has a stability/partition during cell division similar to pSY3. ORF16.2 showed substantial similarity to the initiator replicator protein RepA, which plays an essential role in plasmid replication and is encoded by pGOX3 from *Gluconobacter oxydans* 621H and pCHQ1 from *S. japonicum* UT26S (92% similarity and 86% identity) (Qin and Hartung, 2001; Prust *et al*., 2005) and pALC1 from *Paracoccus alcaliphilus* (81% similarity and 60% identity) (Bartosik *et al*., 2001), respectively, as shown in the pCAR1 plasmid from *Pseudomonas resinovorans* strain CA109 (Shintani *et al*., 2006). An InterProScan search revealed that the pSY2 RepA protein (ORF16.2) contained a putative winged helix-turn-helix motif (RF10134) in its middle section that promotes DNA binding. The phylogenetic tree showed a close relationship of pSY2 RepA to that of pGOX3 and pCHQ1 (Fig. 3). ORF5.2 (66 aa) had 44% identity to the CopG domain protein from *Acidiphilium cryptum* JF-5 (Table 1). CopG, also known as RepA, which is responsible for regulating plasmid copy number, binds to the RepAB promoter and controls synthesis of plasmid replication initiator protein RepB. ORF6.2 (390 aa) was identified as a DNA methylase with high identity (64%) to that of the C-5 cytosine-specific family from *Ruegeria pomeroyi* DSS-3 (Table 1), which protects host DNA against degradation by enzyme restriction. The ORF information associated with DNA replication provides evidence that pSY2 in strain DJ77 is indeed a functionally replicating plasmid.

pSY2 has three putative proteins that seem to be involved in plasmid mobilization for conjugative function. Based on the BLAST results, ORF17.2 (515 aa) showed the highest identity (64%) to the conjugative protein TraG from *Agrobacterium tumefaciens*. ORF1.2 (97 aa) and ORF2.2 (362 aa) were identified as mobilization protein C and a RepB/MobA like protein, which are necessary for self-mobilization, respectively (Table 1).

Analysis of other interesting proteins on pSY2, which have plasmid-like characteristics, further supports that pSY2 could be a functional plasmid in strain DJ77. As shown in Table 1,

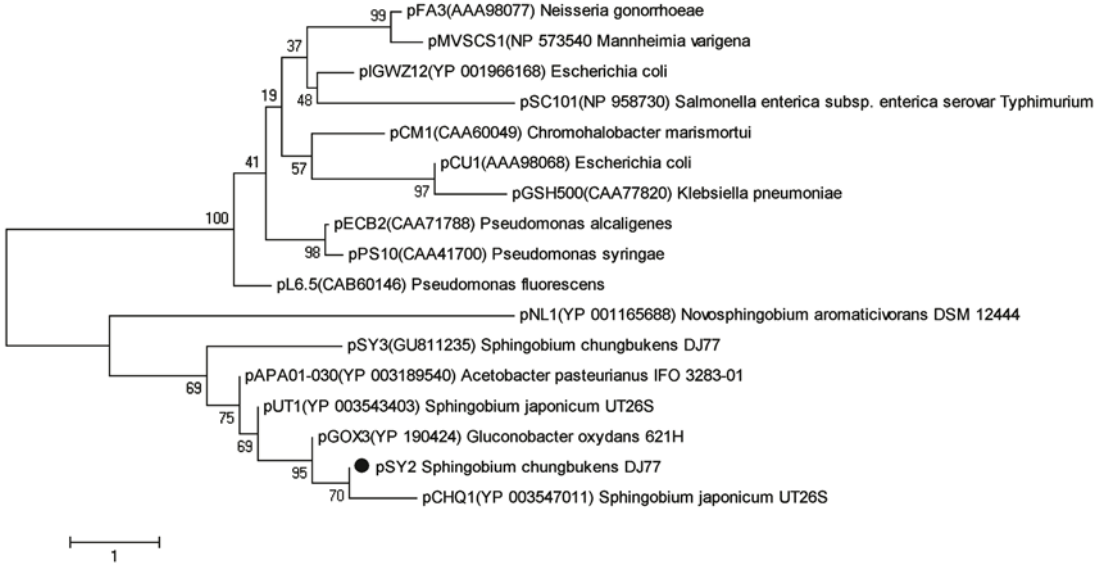

Fig. 3. Phylogenetic tree of open reading frame (ORF) 16.2 (RepA) obtained from alignment with replication initiator proteins. Plasmid names and accession numbers (in parentheses) of the respective proteins are shown. Scale and percentage divergence. The numbers on the branches refer to the percentage confidence, estimated by bootstrap analysis with 1,000 replications.

four ORFs on pSY2 were similar to transposase (*tnp*) genes, suggesting the high possibility of pSY2 gene rearrangement through recombination mediated by *tnp* elements. ORF19.2 (39 aa), ORF20.2 (395 aa), ORF21.2 (408 aa), and ORF22.2 (532 aa) of pSY2 showed identity with genes encoding a transposase of the *Sphingomonas wittichii* RW1 Tn3 family of transposons. This was further supported by the presence of two putative TnpR resolvase genes that code for a site-specific recombinase typical of plasmids. ORF9.2 (183 aa) and ORF18.2 (289 aa) had high similarity (>60% identity) with the resolvase enzyme of *Rhodopseudomonas palustris* GA009 (Table 1).

Other putative functions of pSY2 gene products are transport, fatty acid biosynthesis, stress reactions, and growth rate regulation, ORF3.2 (168 aa) showed 47% identity with the predicted solute carrier organic anion transporter of *Taeniopygia guttata*. ORF8.2 (175 aa) had similarity with the beta-ketoacyl-acyl carrier protein synthase III (FabH) of delta proteobacterium MLMS-1 in fatty acid biosynthesis. ORF12.2 (72 aa) had a 55% identity with the cold-shock DNA domain binding domain protein of *Sphingopyxis alaskensis* RB2256 (Table 1), which functions in various cellular process including low temperature, cellular growth, and nutrient stress. ORF14.2 (282 aa) had a 73% identity function as a universal stress protein, UspA, of *Methylobacterium extorquens* DM4. ORF15.2 (495 aa) could function as a sulphate transporter, and had a 75% identity with that of *Sphingopyxis alaskensis*. In the case of ORF13.2 (113 aa), the putative protein showed high similarity with the transcriptional regulator of the TraR/DksA family, which contributes to growth rate regulation of *Sphingopyxis alaskensis* RB2256. The proteins of these genes may contribute to the biological activities of the plasmid host strain DJ77 for nutrient starvation during active growth and influence metabolism of the priority compounds of human health concern such as PAHs, EPS, and antibiotics in the host strain DJ77.

ORF4.2, ORF7.2, and ORF11.2 showed the highest identities to the hypothetical proteins from various species, such as *S. japonicum* UT26S and *Xylella fastidiosa*. Even though three ORFs on the pSY2 have not been characterized, we cannot rule them out as a non-functional protein in the host.

This work was supported by a research grant from Chungbuk National University in 2009.

References

Azuma, Y., A. Hosoyama, M. Matsutani, N. Furuya, H. Horikawa, T. Harada, H. Hirakawa, and *et al*. 2009. Whole-genome analyses reveal genetic instability of *Acetobacter pasteurianus*. *Nucleic Acids Res.* 37, 5768-5783.

Bartosik, D., M. Witkowska, J. Baj, and M. Wlodarczyk. 2001. Characterization and sequence analysis of the replicator region of the novel plasmid pALC1 from *Paracoccus alcaliphilus*. *Plasmid* 45, 222-226.

Basta, T., A. Keck, J. Klein, and A. Stolz. 2004. Detection and characterization of conjugative degradative plasmids in xenobiotic-degrading *Sphingomonas* strains. *J. Bacteriol.* 186, 3862-3872.

Burenjargal, M., Y.S. Lee, J.M. Yoo, Y.C. Kim, Y.M. Lee, S. Oh, Y.P. Yun, and *et al*. 2007. Endogenous sphingolipid metabolites related to the growth in *Sphingomonas chungbukensis*. *Arch. Pharm. Res.* 30, 317-322.

Ghai, R., T. Hain, and T. Chakraborty. 2004. GenomeViz: visualizing microbial genomes. *BMC Informatics* 5, 198.

Goldberg, S.M., J. Johnson, D. Busam, T. Feldblyum, S. Ferriera, R. Friedman, A. Halpern, and *et al*. 2006. A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proc. Natl. Acad. Sci. USA* 103, 11240-11245.

Kim, S.J., J. Chun, K.S. Bae, and Y.C. Kim. 2000. Polyphasic assignments of an aromatic-degrading *Pseudomonas* sp., strain DJ77, in the genus *Sphingomonas* as *Sphingomonas chungbukensis* sp. nov. *Int. J. Syst. Evol. Microbiol*. 50, 1641-1647.

Koumi, P., H.E. Green, S. Hartley, D. Jordan, S. Lahec, R.J. Livett, K.W. Tsang, and D.M. Ward. 2004. Evaluation and validation of the ABI 3700, ABI 3100, and the MegaBACE 1000 capillary array electrophoresis instruments for use with short tandem repeat microsatellite typing in a forensic environment. *Electrophoresis* 25, 2227-2241.

Nagata, Y., Y. Ohtsubo, R. Endo, N. Ichikawa, A. Ankai, A. Oguchi, S. Fukui, N. Fujita, and M. Tsuda. 2010. Complete genome sequence of the representative γ-hexachlorocyclohexane-degrading bacterium *Sphingobium japonicum* UT26. *J. Bacteriol.* 192, 5852-5853.

Pal, R., S. Bala, M. Dadhwal, M. Kumar, G. Dhingra, O. Prakash, S.R. Prabagaran, and *et al*. 2005. Hexachlorocyclohexane-degrading bacterial strains *Sphingomonas paucimobilis* B90A, UT26 and Sp+, having similar lin genes, represent three distinct species, *Sphingobium indicum* sp. nov., *Sphingobium japonicum* sp. nov. and *Sphingobium francense* sp. nov., and reclassification of *Sphingomonas chungbukensis* as *Sphingobium chungbukense* comb. nov. *Int. J. Syst. Evol. Microbiol.* 55, 1965-1972.

Prust, C., M. Hoffmeister, H. Liesegang, A. Wiezer, W.F. Fricke, A. Ehrenreich, G. Gottschalk, and U. Deppenmeier. 2005. Complete genome sequence of the acetic acid bacterium *Gluconobacter oxydans*. *Nat. Biotechnol.* 23, 195-200.

Qin, X. and J.S. Hartung. 2001. Construction of a shuttle vector and transformation of *Xylella fastidiosa* with plasmid DNA. *Curr. Microbiol*. 43, 158-162.

Rutherford, S.T., J.J. Lemke, C.E. Vrentas, T. Gaal, W. Ross, and R.L. Gourse. 2007. Effects of DksA, GreA, and GreB on transcription initiation: insights into the mechanisms of factors that bind in the secondary channel of RNA polymerase. *J. Mol. Biol*. 366, 1243-1257.

Shin, H.J., S.J. Kim, and Y.C. Kim. 1997. Sequence analysis of the *phnD* gene encoding 2-hydroxymuconic semialdehyde hydrolase in *Pseudomonas* sp. strain DJ77. *Biochem. Biophys. Res. Commun*. 232, 288-291.

Shintani, M., H. Yano, H. Habe, T. Omori, H. Yamane, M. Tsuda, and H. Nojiri. 2006. Characterization of the replication, maintenance, and transfer features of the IncP-7 plasmid pCAR1, which carries genes involved in carbazole and dioxin degradation. *Appl. Environ. Microbiol*. 72, 3206-3216.

Tisa, L.S. and B.P. Rosen. 1990. Transport systems encoded by bacterial plasmids. *J. Bioenerg. Biomembr*. 22, 493-507.

Voss, C. 2007. Production of plasmid DNA for pharmaceutical use. *Biotechnol. Annu. Rev.* 13, 201-222.

Yeon, S.M., B.S. Choi, and Y.C. Kim. 2008. Organization of three rRNA (*rrn*) operons from *Sphingobium chungbukense* DJ77. *J. Microbiol*. 46, 697-703.

Yeon, S.M. and Y.C. Kim. 2009. Characterization of plasmid pSY3 in *Sphingobium chungbukense* DJ77. *J. Microbiol*. 47, 796-800.