

Identification of Strain-Specific Genes in *Rhodococcus erythropolis* Using a Modified HiCEP Method

Yasuo Mitani¹, Tsuneaki Watanabe², Kazuhiro Kondo² and Tomohiro Tamura^{1,3,*}

¹Research Institute of Genome-based Biofactory, National Institute of Advanced Industrial Science and Technology (AIST), Tsukisamu-Higashi, Toyohira-ku, Sapporo 062-8517, Japan; ²AISIN COSMOS R & D CO. LTD., Kazusakamatari, Kisarazu, Chiba, 292-0818, Japan and ³Laboratory of Molecular Environmental Microbiology, Graduate School of Agriculture, Hokkaido University, Kita-9, Nishi-9, Kita-ku, Sapporo 060-8589, Japan

Abstract: The genus *Rhodococcus* exhibits a broad catabolic diversity that is often related to unusual genome diversity. A modified HiCEP method was applied for the identification of strain-specific genes by comparing two different strains. *Rhodococcus erythropolis* JCM3201 and NI86/21 were compared and 11 genes were identified as NI86/21-specific genes.

Key Words: *Rhodococcus*, strain-specific, plasmid, comparative genomics, modified HiCEP.

INTRODUCTION

The genus *Rhodococcus* exhibits a broad range of catabolic diversity [1, 2]. The genome diversity of this genus is exemplified by the existence of unusually diverse catabolic pathways as revealed by the genome sequencing of *Rhodococcus* sp. RHA1 [3]. This strain was also shown to have a potential to produce various secondary metabolites, such as non-ribosomal peptides and polyketides, that are encoded by an unexpectedly high number of corresponding genes [3]. Such genomic diversity has been demonstrated to be derived from the diversity of both chromosome and plasmids, the latter of which are often unusually large [4]. Although in *Rhodococcus* whole genome information is available only for the RHA1 strain, genome diversity among different species or strains has been revealed by the presence of large plasmids that encode various catabolic pathway genes. For example, *Rhodococcus* sp. RHA1 harbors 3 large plasmids [3], *R. erythropolis* BD2 harbors a 210-kbp linear plasmid [5], and *R. erythropolis* PR4 harbors 2 large plasmids in addition to smaller plasmid [6]. Genes encoded on such plasmids are often related to species- or strain-specific functions [1].

Strain-specific genes have been investigated particularly in pathogenic bacteria by utilizing genome information [7]. Comparative genomics have been used for the identification of species- and/or strain-specific genes both by performing whole genome sequencing (e.g., [8]) and by using microarrays (e.g., [9]). Since pathogenicity is often strain specific, and the identification of pathogenic strains is very important in the clinical field, the characterization of microorganisms at a strain level is essential. Further, the identification of pathogenic strain-specific genes will provide insights into the mechanisms of infection. Genome diversity at the strain level is also a characteristic feature of those microorganisms

that live in habitats contaminated by toxic compounds such as polychlorinated biphenyls (PCBs). In addition to *Rhodococcus* sp. RHA1, genome sequencing in *Burkholderia xenovorans* LB400, which is an effective PCB degrader, has also revealed catabolic versatility, including diverse peripheral aromatic pathways [10]. The genome diversity among 3 distinct strains of this organism was demonstrated using genomic microarrays [10].

We conjectured that proteins encoded by strain-specific genes would have important functions, such as catabolic activity of industrial importance or the production of unidentified secondary metabolites. Moreover, a phylogenetic investigation of strain-specific genes will be of assistance in understanding the evolution of *Rhodococcus* genome diversity. In order to obtain strain-specific genes, methods that require genome information can be utilized. These methods are high throughput, which facilitates the handling of many strains. However, genes that are not harbored by the representative strains cannot be detected using these methods. On the other hand, methods that do not require prior knowledge of genome sequences can be used for the identification of such genes. In this report, we therefore took the advantage of the latter approach and aimed to obtain strain-specific genes without performing costly genome sequencing. For this purpose, we applied the modified high coverage expression profiling (HiCEP) method [11] to analyze the difference between *R. erythropolis* strains JCM3201 and NI86/21. By analyzing 23 bands obtained by this method, we could confirm that 11 genes were specific to the NI86/21 strain.

MATERIALS AND METHODOLOGY

Bacterial Strains and Culture Conditions

R. erythropolis strains JCM 3201, NI86/21, and PR4 were obtained from the Japan Collection of Microorganisms, the National Collection of Agricultural and Industrial Microorganisms (Hungary), and the National Institute of Technology and Evaluation (Japan), respectively. The strains were cultured at 30°C in Luria-Bertani (LB) broth (1% Bacto tryptone, 0.5% yeast extract, 1% NaCl). If necessary, cell growth

*Address correspondence to this author at the Research Institute of Genome-based Biofactory, National Institute of Advanced Industrial Science and Technology (AIST), Tsukisamu-Higashi, Toyohira-ku, Sapporo 062-8517, Japan Tel: +81-11-857-8938; Fax: +81-11-857-8980; E-mail: t-tamura@aist.go.jp

was monitored by measuring the optical density at 600 nm (OD_{600}). In the case of plate culture, the medium was solidified by adding 1.5% agar.

Sequencing Analysis

Partial sequences of the 16S rRNA gene, and the *gyrB*, *kasA/B*, and *ltsA* genes were amplified by using the following respective primer pairs; 5'-GACGAACGCTGGCGGCGTGCT-3' and 5'-TTCGTCCCAATCGCCGATCCC-3', 5'-TCCGACTCTACGCCGTCTCC-3' and 5'-AGCACGATCTGTGATAGCGC-3', 5'-CCTGAAGGTTTCACCAGACG-3' and 5'-CCAGAACCGTGAGGACTGAT-3', and 5'-TCGAGCACTACACCGTTCTG-3' and 5'-GCAGTGCGTACTTGGTTGTG-3'. The DNA fragments obtained were directly sequenced from each terminal by using the same primers used for the amplification.

Modified HiCEP Method

A modified HiCEP method was performed as previously described [11]. For the modified HiCEP analysis, each strain was grown to a late log phase ($OD_{600} \sim 0.8$) and was then harvested for RNA isolation. Each band of interest obtained using poly-acrylamide gel electrophoresis was amplified and ligated into pBluescript vector as previously described [11]. After transformation of *Escherichia coli* DH5 α cells, 8 colonies were inoculated and subjected to sequencing analysis. A homology search of sequence data was subsequently performed using BLAST at the National Center for Biotechnology Information.

Polymerase Chain Reaction

Polymerase chain reaction (PCR) analysis for the validation of isolated fragments was performed as follows. A 25- μ l reaction volume containing 0.1 μ g cDNA, which was prepared for the modified HiCEP analysis, was amplified using KOD plus polymerase (Toyobo). The PCR cycle conditions were 94°C for 2 min following 25 cycles of 94°C for 15 s, 58°C for 30 s, and 68°C for 30 s or 1 min (depending on the length of target size), then 68°C for 2 min. Specific primers for each fragment were designed using Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi).

Reverse Transcriptase PCR

Reverse transcriptase PCR (RT-PCR) was performed using a ThermoScript RT-PCR kit (Invitrogen) according to the manufacturer's manual. One microgram of RNA was subjected to reverse transcription in a volume of 20 μ l at 55°C. An aliquot of the reaction product was used as a template for the following PCR. The PCR was performed as described in the previous section.

RESULTS AND DISCUSSION

Nucleotide Sequence Diversity Among *R. erythropolis* Strains

The modified HiCEP method uses 2 different restriction enzymes to create gene fragments that represent the population of each transcript. In order to identify strain-specific genes using this method, nucleotide substitutions should be taken into consideration. If there is a substitution in the restriction enzyme site or adjacent two bases used for the

analysis, a single gene will produce different bands among different strains. These bands can thereby be detected as strain-specific bands. In order to investigate the nucleotide substitution among strains from the same species, we sequenced short regions of 4 genes. The 16S rRNA sequence is known to be conserved within the same species and is also used to classify organisms at a species level [12]. We investigated the 16S rRNA sequences from the strains JCM3201, NI86/21, and PR4 of *R. erythropolis* and found that the sequences were identical in the region analyzed. *gyrB* is also a conserved housekeeping gene. This gene is, however, more diverse than 16S rRNA and can be used for classification at a strain-level resolution [13]. We compared the sequence of this gene among the aforementioned strains. By sequencing approximately 1 kbp regions in the gene, 6 nucleotide substitutions were found between JCM3201 and NI86/21, and 6 nucleotide substitutions were found between JCM3201 and PR4. We also analyzed the sequences of the nonessential genes *ltsA* and *kasA/B*. By sequencing approximately 1 kbp regions in the genes, 7 and 1 nucleotide substitutions were found in *ltsA* and *kasA/B*, respectively between JCM3201 and NI86/21. In the comparison between JCM3201 and PR4, 11 and 5 substitutions were found in *ltsA* and *kasA/B*, respectively. We also sequenced other strains and found that the substitution number between JCM3201 and NI86/21 was lower than that in any other pair of strains analyzed. Therefore, we used these 2 strains for the comparative analysis. In this sequencing analysis, the 25 substitutions for an approximate 1 kbp region of a certain ORF were the highest counted (data not shown).

Identification of Genes Specific to *R. erythropolis* NI86/21

In order to determine if the modified HiCEP method can be applied to a comparative genomic study of different strains from the same species, *R. erythropolis* NI86/21 was analyzed using *R. erythropolis* JCM3201 as a reference strain. *R. erythropolis* NI86/21 is a herbicide-degrading strain with biosafener activity [14]. *R. erythropolis* JCM3201 is one of the type strains of *R. erythropolis*. In the modified HiCEP method, 2 different restriction enzymes are used to obtain DNA fragments that theoretically reflect the original mRNA population [11]. Initially, we used *Pst*I and *Pvu*I for the digestion. Selective PCR was then performed using 16 combinations of primers corresponding to all variations of 3' terminal nucleotides. Here, we present only a representative result of the electrophoresis using 8 combinations (Fig. 1). The band patterns of these strains appeared similar. In the case of other primer combinations, the band patterns between the strains were also similar (data not shown). By comparing the band patterns, we found over 100 bands that were detectable in either strain. However, some of these were difficult to cut out from the gel due to the following reasons: the target band located amongst bands of similar size and/or the target band size was too small to be separated well on the gel used. In this study, we selected 18 bands that were found only in the NI86/21 sample in order to simplify the subsequent handling and analysis. Each band was analyzed by sequencing after the cloning of the bands; 8 clones were subjected to sequencing analysis in order to identify the major fragment. This identification is important because bands that are obtained using the modified HiCEP method usually contain a mixture of different DNA fragments, as

previously reported [11]. The results of the sequencing analysis of each band are summarized in Table 1. In order to obtain further bands that are specific to NI86/21, we also used a *Pst*I and *Xho*I combination for the digestion. In this case, 5 bands that were detected only in the NI86/21 sample were analyzed. The results of the sequencing analysis are summarized in Table 1. In order to examine if these gene fragments exist in the genome of strain NI86/21 but not in JCM3201, we performed PCR analysis using total DNA extracted from each strain as a template. As a result of the PCR analysis, 11 fragments were amplified only when using total DNA from the NI86/21 strain (Table 1).

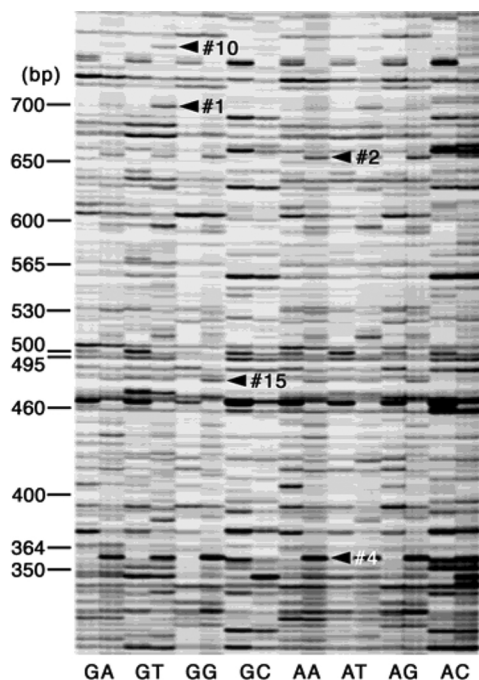


Fig. (1). Band patterns obtained by the modified HiCEP analysis of *R. erythropolis* strains JCM3201 and NI86/21. Bands were separated by using a 6% poly acrylamide gel. Eight combinations of primers are indicated below the gel image (from GA to AC). For each primer combination, the NI86/21 sample and the JCM3201 sample were loaded on the left side and right side, respectively. In this gel image, 5 bands (#1, #2, #4, #10, and #15) that were excised are indicated by arrowheads.

Six out of 11 fragments analyzed showed similarity to putative non-ribosomal peptide synthetase. These genes might encode strain-specific secondary metabolite because many actinomycetes produce strain-specific non-ribosomal peptides as secondary metabolites [15]. Most of other genes except for a gene included in band 2 showed similarity to hypothetical proteins. Band 24 showed similarity to a hypothetical protein encoded in one of the large plasmid of *R. erythropolis* PR4, suggesting a possibility that NI86/21 strain have a plasmid partially similar to that of PR4 strain. In the case of band Xh5, most region of the gene fragment showed similarity to hypothetical protein RHA1_ro08522. However, its 3' end showed similarity to glutamine transport ATP-binding protein GlnQ, which is located next to a gene encoding hypothetical protein RHA1_ro05666. This means that these two genes are separately located in the chromosome of RHA1 strain while they are next to each other in

NI86/21 strain, suggesting this region is not conserved between NI86/21 and RHA1 strains. Although the sequence of this region in JCM3201 strain is not available, it might be different between JCM3201 and NI86/21 strains.

Based on an analysis of 23 bands, we identified 11 gene fragments that were specific to the NI86/21 strain. Nevertheless, the number of NI86/21-specific genes is not limited to this result. The modified HiCEP analysis monitors expressed genes but not the genome itself. Therefore, we were unable to detect the genes that were not expressing under the culture condition used in this analysis. Furthermore, in this study we observed over 100 bands that were specific to either strain; however, due to technical reasons, we analyzed only 23 bands.

The HiCEP method is a technique used to monitor gene expression based on the quantity of mRNA. Therefore, if a gene from either strain is expressing at a very low level, the modified HiCEP method may detect such a gene as a strain-specific band. In order to examine the expression level of each fragment, we performed RT-PCR analysis. However, when the RT-PCR products were analyzed by ethidium bromide staining, there appeared no expression differences between the strains (data not shown). These results suggest that nucleotide substitution is a more critical factor than the difference in gene expression levels.

We used PCR amplification for the confirmation of the modified HiCEP analysis because using PCR it is relatively easy to analyze many fragments. However, in some cases the results were not clear. There were several fragments obtained from a single band that could not be amplified by PCR using either strain (e.g., band 2 and 8). In order to confirm the existence of the undetected genes, it will be necessary to perform other experiments using different sets of primers or to perform Southern hybridization. In the modified HiCEP method, a single band usually contains several gene fragments. In many cases, a fragment of highest numbers counted was assumed to represent the target band. However, it should be noted that in band 2 and 19 the fragments confirmed to be strain specific were not fragments of highest numbers counted. In the case of band 5, a fragment was identified as JCM3201-specific by the PCR analysis. In these cases, further analysis, such as Southern hybridization, may be needed.

Comparison between PR4 and JCM3201

We next compared the band patterns from strains PR4 and JCM3201. We used the *Pst*I and *Pvu*I combination and analyzed the band pattern. However, the band patterns from these 2 strains were apparently different when compared with the pattern difference observed between strains NI86/21 and JCM3201 (data not shown); this disparity was expected based on the results of nucleotide substitution analysis. Despite the pattern difference, we identified 2 bands that were specific to PR4 by analyzing 19 bands that were detected only in the PR4 lanes (data not shown). Both of these bands were derived from a large plasmid, pREL1.

The apparent difference in band patterns between strains JCM3201 and PR4 may be due to the existence of different large plasmids in these strains. We obtained gene fragments derived from pREL1, suggesting that these genes exist only in the PR4 strain. On the other hand, we detected a large

Table 1. NI86/21-Specific Genes

Band ^a	N ^b	Length (bp)	Homologue ^c	Identity ^d (%)	PCR ^e	
					JCM3201	NI86/21
1	7	653	putative non-ribosomal peptide synthetase ^{*1}	57	-	+
2	4	613	hypothetical protein RHA1_ro05779	75	-	-
	1	612	transcriptional regulator, LysR family ^{*2}	47	-	+
	1	613	conserved hypothetical protein ^{*3}	27	-	-
3	2	340	non-ribosomal peptide synthetase	51	-	+
	1	344	none		-	-
	1	350	glycine hydroxymethyltransferase	97	+	+
4	4	316	RNase III	34	+	+
5	3	810	lysine tRNA ligase	58	+	+
	3	812	probable multisubunit Na ⁺ :H ⁺ antiporter MnhE subunit	88	+	+
	1	809	ABC drug resistance transporter, ATP-binding protein	67	+	-
6	8	1008	translation initiation factor IF-2 ^{*4}	67	-	-
8	6	783	aspartate ammonia-lyase	93	+	+
	1	780	putative IcmB-like type IV secretion system protein ^{*5}	35	-	-
	1	780	hypothetical protein CE0887 ^{*3}	21	+	+
9	7	866	hypothetical protein RHA1_ro01299	69	+	+
	1	866	probable L-malate dehydrogenase ^{*6}	36	-	-
10	7	706	non-ribosomal peptide synthetase	53	-	+
	1	702	transcriptional regulator, GntR family ^{*7}	62	-	-
13	2	481	chromosome partition protein	72	+	+
	1	484	polyketide synthase	66	+	+
14	1	474	probable fatty-acyl-CoA synthase	87	+	+
	1	477	DNA polymerase I ^{*8}	99	+	+
15	1	436	non-ribosomal peptide synthetase	36	-	+
18	4	340	non-ribosomal peptide synthetase	51	-	+
	2	343	hypothetical protein RHA1_ro00975	70	+	+
	1	344	none		+	+
19	3	320	uroporphyrinogen decarboxylase	90	+	+
	2	325	phosphoglucosamine mutase	64	+	+
	1	324	none		-	+
	1	328	sigma factor	54	+	+
	1	322	conserved hypothetical protein ^{*9}	54	+	+
21	3	291	hypothetical protein PaerPA_01002889 ^{*6}	36	-	+
	3	292	hypothetical protein RHA1_ro04972	95	+	+
	1	287	probable fatty-acyl-CoA synthase	79	+	+

(Table 1) contd....

Band ^a	N ^b	Length (bp)	Homologue ^c	Identity ^d (%)	PCR ^e	
					JCM3201	NI86/21
22	2	257	phenylalanine-tRNA ligase beta subunit	82	+	+
	1	253	possible adenylate cyclase	85	+	+
	1	242	methionine synthase	86	+	+
	1	254	hypothetical protein RHA1_ro02104	80	+	+
23	7	257	phenylalanine-tRNA ligase beta subunit	82	+	+
	1	254	hypothetical protein RHA1_ro02104	80	+	+
24	2	235	none		+	+
	1	232	hypothetical protein pREL1_0234 ^{*10}	72	-	+
	1	228	porphobilinogen synthase	94	+	+
	1	233	hypothetical protein RHA1_ro01965	50	+	+
	1	234	hypothetical protein RHA1_ro02110	83	+	+
Xh2	5	999	putative non-ribosomal peptide synthetase ^{*1}	72	-	+
	1	997	transaldolase	75	+	+
Xh4	8	851	polyketide synthase	61	+	+
Xh5	4	700	possible dipeptide-binding protein	31	+	+
	3	690	glutamine transport ATP-binding protein GlnQ	24	-	+
Xh9	3	461	putative galactofuranosyltransferase ^{*1}	85	+	+
Xh10	3	338	putative <i>N</i> -acetylmuramoyl-L-alanine amidase	81	+	+
	2	334	CTP synthetase	87	+	+
	1	327	acetamidase/formamidase ^{*11}	85	+	+

^aBands obtained using a *Pst*I and *Pvu*I combination are shown by number only. Bands obtained using a *Pst*I and *Xho*I combination are designated by an Xh plus number.

^bThe number of clones identified by sequencing 8 clones for each band. The total number for each band is not always 8 because the clones that had sequencing errors are excluded from the list. Fragments for which the size is apparently shorter or longer than the size estimated by the modified HiCEP analysis were excluded from the list.

^cOrganisms of the homology search results were *Rhodococcus* sp. RHA1 if not indicated by the asterisk: ^{*1}*Nocardia farcinica*; ^{*2}*Pseudomonas mendocina*; ^{*3}*Corynebacterium efficiens*; ^{*4}*Janibacter* sp. HTCC2649; ^{*5}*Xanthomonas campestris*; ^{*6}*Frankia* sp. EAN1pec; ^{*7}*Pseudomonas aeruginosa*; ^{*8}*Pseudomonas fluorescens*; ^{*9}*Rhodococcus* sp. ATCC 15963; ^{*10}*Vibrio cholerae*; ^{*11}*R. erythropolis* PR4; ^{*12}*Mycobacterium* sp. JLS

^dAmino acid identity between each ORF predicted for the gene fragment and a homologue identified by BLAST search.

^ePCR results using each genome as a template. +, amplified; -, not amplified.

linear plasmid (approximately 390 kbp) in the JCM3201 strain using pulse-field gel electrophoresis (unpublished data).

Computational Analysis of Restriction Enzymes Combination

It was very difficult to select an appropriate combination of restriction enzymes for the modified HiCEP analysis. In order to make a rational selection of restriction enzymes, we decided to perform a computational analysis. For this purpose, we used *R. erythropolis* PR4 because the sequences of the large plasmids harbored by this strain have recently been determined [6]. This strain has 2 large plasmids; one of these is a 272-kbp linear plasmid named pREL1 and the other is a 104-kbp circular plasmid named pREC1. Therefore, we considered that this strain could be used for a prediction of the gene fragments that potentially generated using the modified

HiCEP method. For an exact prediction, we needed transcript annotations including untranslated regions, as well as those of open reading frames (ORFs). However, the sequence information only provides ORF annotation. Therefore, we assumed an ORF as a transcript for the fragment prediction. In this study, the apparent overlap of ORFs was assumed to be an operon and was analyzed as a single transcript, while ORFs with only a few nucleotide intervals were assumed to be independent transcripts. On the basis of these assumptions, we predicted that analysis of all 400 ORFs on the 2 large plasmids would yield a total of 318 transcripts.

In order to achieve extensive genome sequence coverage, it is preferable to use a combination of restriction enzymes that produces a large number of bands. However, the band sizes will be shorter when using such a combination and this may give rise to practical problems, particularly if the band size is less than 200 bp. Shorter bands usually include large

numbers of DNA fragments due to the limitation of gel resolution; this give rise to problems in the identification of positive fragments. In the comparison between strains JCM3201 and NI86/21, we selected the *Pst*I and *Pvu*I combination based on observations from several preliminary experiments. From a computational analysis of PR4 plasmids, 23 fragments of over 200 bp were predicted for this combination. We performed a computational analysis to find other combinations that produced more bands of over 200 bp than produced by the *Pst*I and *Pvu*I combination. By analyzing over 1400 combinations, we found several combinations that produce more bands. For example, the *Sal*I and *Eag*I combination was predicted to produce 36 bands of over 200 bp. These combinations warrant further investigation.

By using computational analysis, it was demonstrated that *Pst*I and *Pvu*I combination covered only 7% of the genes encoded on large plasmids when the band size was limited to that greater than 200 bp. On the other hand, using the *Sal*I and *Eag*I combination, the coverage was estimated to increase to 11%. There is a possibility that greater coverage may be achieved by performing analysis using a few appropriate combinations of restriction enzymes; for example, the coverage was estimated to increase to 25% when using 3 combinations of restriction enzymes.

In this study, we applied a modified HiCEP method for the identification of strain-specific genes. In the case of strains that exhibit low number of nucleotide substitutions between them, the method is very useful for the purpose. To date, strain-specific genes have been investigated particularly in pathogens due to a clinical importance. In non-pathogenic bacteria, however, strain-specific genes may be good candidates for the production of useful proteins of industrial and/or environmental importance. As demonstrated here, the modified HiCEP method can be used to identify strain-specific genes, particularly those from strains for which genome information is currently not available.

ACKNOWLEDGEMENTS

We thank all the members of our laboratory for their technical help and for the invaluable discussions.

REFERENCES

- [1] van der Geize R, Dijkhuizen L. Harnessing the catabolic diversity of rhodococci for environmental and biotechnological applications. *Curr Opin Microbiol* 2004; 7(3): 255-61.
- [2] Larkin MJ, Kulakov LA, Allen CCR. Biodegradation and Rhodococcus – masters of catabolic versatility. *Curr Opin Biotechnol* 2005; 16(3): 282-90.
- [3] McLeod MP, Warren RL, Hsiao WWL, *et al.* The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci* 2006; 103(42): 15582-7.
- [4] Gürtler V, Mayall BC, Seviour R. Can whole genome analysis refine the taxonomy of the genus *Rhodococcus*? *FEMS Microbiol Rev* 2004; 28(3): 377-403.
- [5] Stecker C, Johann A, Herzberg C, Averbhoff B, Gottschalk G. Complete nucleotide sequence and genetic organization of the 210-kilobase linear plasmid of *Rhodococcus erythropolis* BD2. *J Bacteriol* 2003; 185(17): 5269-74.
- [6] Sekine M, Tanikawa S, Omata S, *et al.* Sequence analysis of three plasmids harboured in *Rhodococcus erythropolis* strain PR4. *Environ Microbiol* 2006; 8(2): 334-46.
- [7] Raskin DM, Seshadri R, Pukatzki SU, Mekalanos JJ. Bacterial genomics and pathogen evolution. *Cell* 2006; 124(4): 703-14.
- [8] Welch RA, Burland V, Plunkett III G, *et al.* Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci* 2002; 99(26): 17020-4.
- [9] Dziejman M, Serruto D, Tam VC, *et al.* Genomic characterization of non-O1, non-O139 *Vibrio cholerae* reveals genes for a type III secretion system. *Proc Natl Acad Sci* 2005; 102(9): 3465-70.
- [10] Chain PS, Denev VJ, Konstantinidis KT, *et al.* *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *Proc Natl Acad Sci* 2006; 103(42): 15280-7.
- [11] Mitani Y, Suzuki K, Kondo K, Okumura K, Tamura T. Gene expression analysis using a modified HiCEP method applicable to prokaryotes: a study of the response of *Rhodococcus* to isoniazid and ethambutol. *J Biotechnol* 2006; 123(3): 259-72.
- [12] Savelkoul PHM, Aarts HJM, de Haas J, *et al.* Amplified-fragment length polymorphism analysis: the state of an art. *J Clin Microbiol* 1999; 37(10): 3083-91.
- [13] Huang WM. Bacterial diversity based on type II DNA topoisomerase genes. *Annu Rev Genet* 1996; 30: 79-107.
- [14] Nagy I, Schoofs G, Compernelle F, Proost P, Vanderleyden J, De Mot R. Degradation of the thiocarbamate herbicide EPTC (S-ethyl dipropylcarbamothioate) and biosafening by *Rhodococcus* sp. strain NI86/21 involve an inducible cytochrome P-450 system and aldehyde dehydrogenase. *J Bacteriol* 1995; 177(3): 676-87.
- [15] Ayuso A, Clark D, Gonzalez I, Salazar O, Anderson A, Genilloud O. A novel actinomycete strain de-replication approach based on the diversity of polyketide synthase and nonribosomal peptide synthetase biosynthetic pathways. *Appl Microbiol Biotechnol* 2005; 67(6): 795-806.