[SII-2]

# A Commensal Thermophile, *Symbiobacterium toebii*: Distribution, Characterization, and Genome Analysis

Jin-Woo Bae[1], Kwang Kim[1], Jae Jun Song[1], Jae Seok Ha[1], Joong-Jae Kim[1], Gwan-Tae Kang[1], Mi-Hwa Kim[1], Seung-Pyo Hong[2], Moon-Hee Sung[1,2]*

[1]Microbial Genomics Lab., Korea Research Institute of Bioscience and Biotechnology (KRIBB), Yusong, Daejeon, Korea, [2]BioLeaders Corp. Bio-Venture Center, KRIBB, Yusong, Daejeon, Korea

## Abstract

A commensal thermophile, *Symbiobacterium toebii*, isolated from hay compost (toebii) in Korea commensally interacted with a thermophilic *Geobacillus toebii* sp. nov., which was a new species within the genus *Geobacillus* on the basis of the phenotypic traits and molecular systematic data. *S. toebii* required the crude extracts and/or culture supernatant of the *Geobacillus toebii* for axenic growth and could grow on the temperature between 45 and 70℃ (optimum: 60℃; 2.4 h doubling time) and pH 6.0 and 9.0 (optimum: pH 7.5). The G+C content of the genomic DNA was 65 mol%, and the major quinones were MK-6 and MK-7. A phylogenetic analysis of its 16S rDNA sequence indicated that *Symbiobacterium toebii* was closely related with solely reported *Symbiobacterium thermophilum*. The presence of the commensal thermophile 16S rDNA and accumulation of indole in all the enriched cultures indicate that *Symbiobacterium toebii* is widely distributed in the various soils. The genome of *S. toebii* constituted a circular chromosome of 3,280,275 base pairs and there was not an extra-chromosomal element (ECE). It contained about 4,107 predicted coding sequences. Of these protein coding genes, about 45.6% was encoded well-known proteins and annotated the functional assignment of 1,874 open reading frames (ORFs), and the rest predicted to have unknown functions. The genes encoding thermostable tyrosine phenol-lyase and tryptophan indole-lyase were cloned from the genomic DNA of *S. toebii* and the enzymatic production of L-tyrosine and L-tryptophan was carried out with two thermostable enzymes overexpressed in recombinant *E. coli*.

## Introduction

There are large numbers of unidentified microorganisms in natural environments that could be seen only in microscope or detected by molecular ecological methods. Although large numbers of new thermophilic bacteria have been isolated in recent years, only a small fraction of environmental bacterial communities can be cultivated by current techniques because of our inability to understand and reproduce the micro-environmental habitats in nature. It has been assumed that biological interactions (e.g., symbiotic

relationships) are essential factors for the growth of some microorganisms. For example, the growth of bacterial parasite, *Bdellovibrio*, requires a Gram-negative bacteria host. Syntrophic bacteria can only grow when methanogens remove hydrogen, the metabolic product of the syntrophic bacteria.

A new thermophilic bacterium, strain SC-1 detected from compost during the screening of thermostable tyrosine phenol-lyase-producing bacteria exhibited a commensal interaction with a thermophilic *Geobacillus* strain (Lee et al. 1996; 1997). So far, the growth of strain SC-1 has never been observed in the absence of *Geobacillus* sp. strain SK-1 in enrichment cultures. The strain SC-1 has been isolated using a medium containing the crude extracts and culture supernatant of strain SK-1 (Rhee et al. 2000). This type of bacterial interaction has been also reported previously by Suzuki et al. (1988). Recently, with the use of a dialyzing culture vessel, the independent growth of *Symbiobacterium thermophilum* was confirmed and *S. thermophilum* validly described for the first time (Ohno et al. 1999, 2000). However, the more researches were needed for evolutionary phylogeny of these microorganisms and understanding of their microbial interactions in ecosystem.

Since the Department of Energy (DOE) initiated microbial genome project in 1994, genomes of about 30 microorganisms were completely sequenced by several centers such as TIGR, Diversa Co., LION Bioscience, JGI, Sanger Centre, and several universities. The fact that most of these microbes were extremophiles and pathogenic microorganisms indicated that the genomic information of these microbes could be used practically in industrial and medical fields. For example, *Methanococcus jannaschii* isolated from a sample, which was collected at a base of deep-sea thermal vent in Pacific Ocean, had the ability to produce methane and this feature has made us to exploit material resource for new form of fuel. In 1996, completely sequenced genome of *M. jannaschii* indicated that about 65% of predicted ORFs encoding proteins was revealed to be unknown genes (Carol et al. 1996). Many scientists have thought it to be exciting area for future study and profound resources for new enzymes which could be applicable in industrial field.

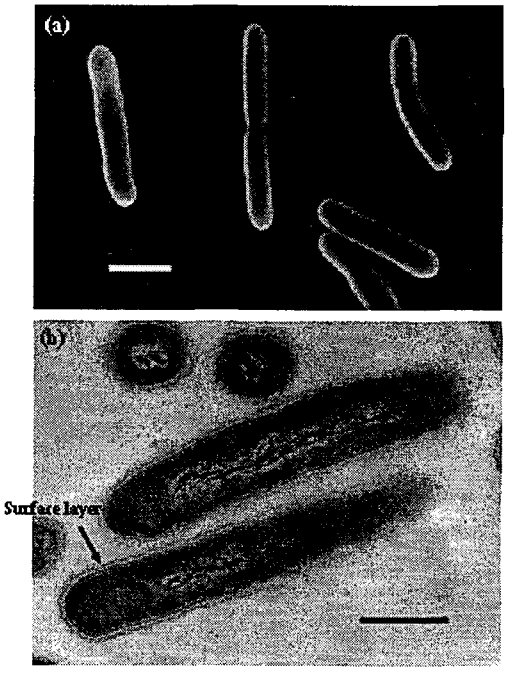


Fig. 1. Electron micrographs of strain SC-1. (a) Scanning electron micrograph of exponentially growing cells occurring singly or in pairs; bar, 0.5 μm. (b) Electron micrograph of a thin section of an exponentially growing cells; bar, 0.3 μm.

The aim of this work was to characterize the isolated SC-1 from compost in Korea, which exhibits a commensal interaction with a *Geobacillus* sp. strain SK-1. Accordingly, in this paper the morphological and physiological characteristics of the isolate SC-1 are described along with its phylogenetic position based on 16S rDNA sequences. We also carried out complete sequencing of genome of *S. toebii* since its thermostable enzymes have many applications at industrial field.

## Results & Discussion

### Enrichment and isolation

Using the enrichment techniques described above, a mixed culture containing strain SC-1 and *Geobacillus* sp. strain SK-1 was obtained. Although the growth of strain SC-1 was never observed in the absence of *Geobacillus* sp. strain SK-1, the growth of the strain SC-1 did not require live *Geobacillus* strain SK-1 when extract and/or culture supernatant of strain SK-1 were present in the medium. Therefore, strain SC-1 was isolated using an MBM agar containing the crude extract and/or culture supernatants of strain SK-1 under an anoxic atmosphere. After 3 days of incubation under microaerobic or anoxic conditions, the colonies observed on the MBM agar plates were circular, transparent, and less than 0.1 mm in diameter.

### Morphological and physiological characteristics

The cells of strain SC-1 were found to be nonmotile, slightly curved rod, 1-5 μm long, and 0.2-0.3 μm wide (Fig. 1a). No flagella were observed by transmission microscope of negatively stained cells. The cells of strain SC-1 stained gram-negative and no spore formation on the sporulation agar medium or MBM was observed. A thin section electron microscopy is shown in Fig. 1b.

The growth of strain SC-1 was detected under both microaerobic and anoxic conditions, since a pure culture of strain SC-1 grew optimally under a reduced oxygen tension. Under anoxic conditions, strain SC-1 grew anaerobically reducing nitrate to nitrite stoichiometrically. Therefore, determination of accumulated nitrite was used as an alternative for growth measurement of strain SC-1. The organism grew at temperatures between 45 and 70℃, with an optimum of 60℃, and the generation time at this temperature was about 2.4 h at pH 7.5 (Fig. 2a). No growth was observed at 80℃. The growth of the new isolate at 60℃ occurred at pH values of 6.0-9.0, with an optimum about 7.5 (Fig. 2b). No growth was detected at pH values below 6.0 and above 9.0.

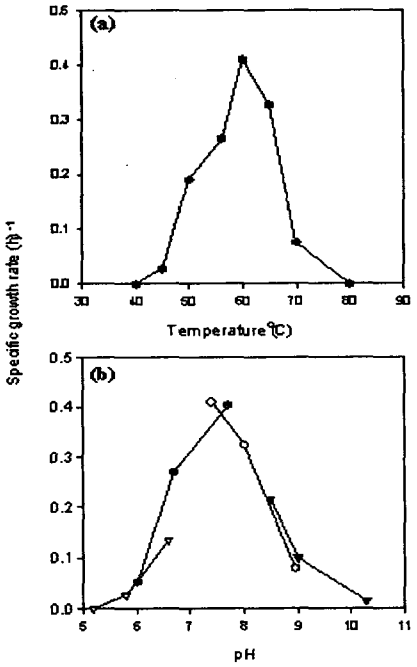


Fig. 2. Effects of temperature (a) and pH (b) on growth of strain SC-1 (a) Growth was determined in MBM at pH 7.5. No growth occurred below 45°C and above 70°C. (b) Growth was determined in MBM with various buffers at 60°C. The pH of the medium containing appropriate buffer systems was adjusted with HCl and NaOH at room temperature. Symbols: ▽, citrate-$Na_2HPO_4$ buffer; ●, phosphate buffer; ○, Tris-HCl buffer; ▼ glycine-NaOH buffer.

### Biochemical characteristics

It was reported that heat stable factors and unstable factor could support the growth of SC-1 in a previous paper (Rhee et al. 2000), therefore, this study attempted to purify and identify the commensal factors from its commensal

partner *Geobacillus* sp. SK-1. It was shown that heat stable (low molecular weight) factors were always not essential for the growth of strain SC-1. The heat unstable factors were aggregated and precipitated at pH 4.0 and renatured by returning to pH 7.0. The resulting factor was irreversibly inactivated by phenol or protease treatment, thereby suggesting that the heat unstable factors may be proteins. Since the membrane fraction of *Geobacillus* showed low growth activity (less than 5 % of cytoplasmic fraction), the factors may be in the cytoplasm. The more researches for the function of the factors should be required to understand the unknown microbial interactions in the ecosystem

Strain SC-1 showed about 5 times less growth yield in pure cultures than in mixed cultures. In addition because of its dependency on the extract and culture supernatant of the strain SK-1 and its low growth yield in the pure culture, it was difficult to determine conventional phenotypic characteristics, such as nutrient utilization, growth at different media, and acid production from sugars. Tryptophan indole-lyase (0.34 U/mg-protein) and tyrosine phenol-lyase (0.01 U/mg-protein) were induced by tryptophan and tyrosine, respectively. However, tyrosine and tryptophan had no effect on the growth of strain SC-1.

## Chemotaxonomic characteristics and DNA base composition

The major quinones found in strain SC-1 were 61% menaquinone 6 (MK-6) and 39% menaquinone 7 (MK-7). The fatty acid profile of strain SC-1 was characterized by having mainly 39% iso-$C_{15:0}$, 28% iso-$C_{17:0}$, 10% iso-$C_{16:0}$, 7% $C_{16:0}$, 7% anteiso-$C_{17:0}$, 2% anteiso-$C_{15:0}$ and 2% $C_{18:0}$. Strain SC-1 did not contain diaminopimelic acid as the diamino acid in its cell wall. The G+C content of strain SC-1 was 65 mol%.

## 16S rDNA sequence analysis

An almost complete 16S rDNA sequence of strain SC-1, comprising 1521 nucleotides (>96% of *E. coli* 16 rDNA sequence), was determined in this study. An inspection of the predicted secondary structures and evaluation by the CHECK-CHIMERA program of RDP indicated that the rDNA sequence of SC-1 was free of artifacts. A BLAST search was conducted to obtain those sequences that were potentially related to that of strain SC-1. The 16S rDNA sequence of strain SC-1 showed about 98 % sequence similarity with that of *Symbiobacterium thermophilum* (AB004913) reported solely until now. These result indicated that strain SC-1 is a member of *Symbiobacterium* subphylum. However, based on a sequence similarity analysis, the 16S rDNA sequence of the isolate exhibited less than 87% similarity with all other known sequences. The most similar 16S rDNA sequences were those of *Geobacillus thermoleovorans, Alicyclobacillus acidocaldarius, Moorella thermoautotrophicum, Moorella thermoaceticum, Thermaerobacter marianensis, Desulfotomaculum austalicum, Thermoanaerobacter ethanolicus,* and *Thermoanaerobacter thermocopriae* with 85-87% similarities except *S. thermophilum.* As such, the 16S rDNA sequence analysis indicated that *Symbiobacterium* is closely related to the *Bacillus-Clostridium* subphylum of Gram-positive bacteria, even though the isolate appears to be phylogenetically distant from these bacteria at the genus level.

A phylogenetic analysis by the NJ and ML methods suggested a similar topology for the affiliation of the new isolate, indicating a distinct branch denoting an early divergence from the genera of the *Bacillus-Clostridium* subphylum. Although the base composition disparities between the genomic DNAs of low-G+C-content gram-positives range from 31.0 to 72.5%, the base composition disparities between the 16S rDNA sequences were relatively small (from 51.7 mol% G+C content in *Paenibacillus alvei* to 63.0 mol% G+C content in *Thermaerobacter marianensis*). The 16S rDNA of strain SC-1 had a 59.0 mol% G+C content, which is similar to that of thermophilic *Geobacillus* species (from 54.1 mol% in *Geobacillus thermoglucosidasius* to 58.7 mol% in *Geobacillus thermoleovorans*). Furthermore, when using a transversion analysis for the generation of a phylogenetic tree, it was confirmed that the base composition disparities had little influence on the tree topology.

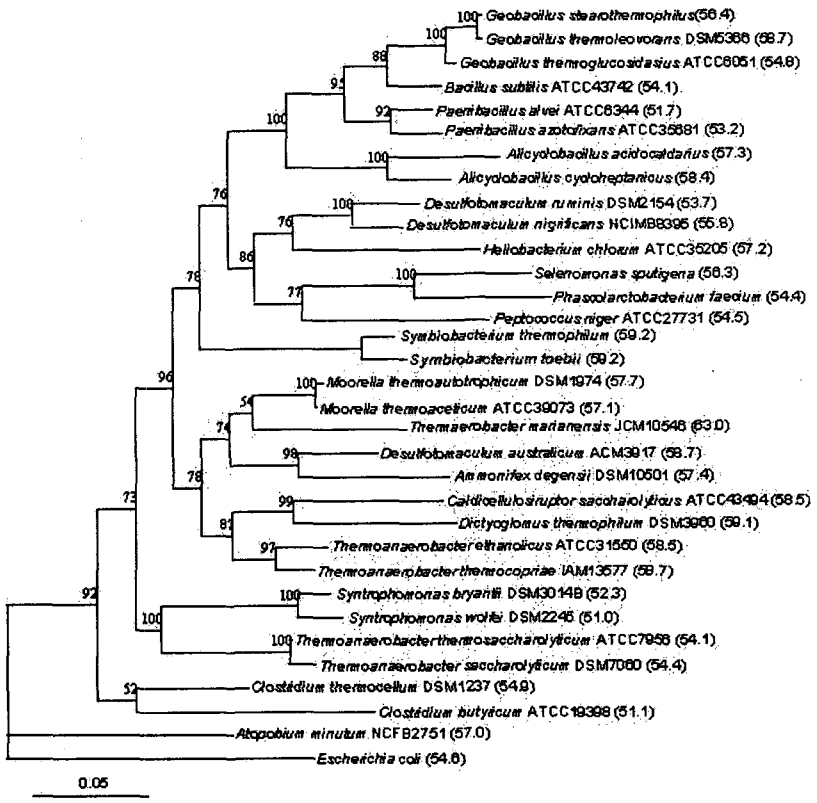


Fig. 3. Phylogenetic dendrogram of representative low G+C-content, spore-forming bacteria within the *Bacillus-Clostridium* subphylum based on 16S rDNA sequences by using the ML method of 1213 homologous positions of sequence. Numbers on tree represent the bootstrap values (expressed as a percentage of 1000 replications). Numbers after the name are the G+C content in mol% of homologous positions of sequence. Scale bar indicates 5 substitutions per 100 nucleotides. The *E. coli* 16S rDNA sequence served as an outgroup.

## DNA-DNA hybridization

For the elucidation of novelty of new isolate, strain SC-1, the DNA-DNA hybridization experiments were carried out. The test showed that chromosomal homology between *Symbiobacterium thermophilum* and *Symbiobacterium* sp. SC-1 is approximately 30% indicating that the new isolate was novel species that is a member of *Symbiobacterium* genus.

## *Symbiobacterium toebii* sp. nov., a new commensal thermophile isolated from compost

Suzuki et al. (1988) described a commensal bacterium "*Symbiobacterium thermophilum*" with physiological similarities to strain SC-1; growth temperature, optimal pH, and tyrosine phenol-lyase and

tryptophan indole-lyase activity, etc. Recently, Ohno et al. (1999, 2000) established a pure culture of *S. thermophilum* from its supporting *Geobacillus* strain, validly described as *S. thermophilum*. Since strain SC-1 exhibits a high similarity to *S. thermophilum* in its 16S rDNA and physiological characteristics, it was tentatively identified as *Symbiobacterium* sp. strain SC-1. However, *S. thermophilum* does not have round-end rod morphology and its cell wall structure is also slightly different and the genome size by pulsed-filed gel was different from that of *S. thermophilum* (Hong et al. 2000). A thin-sectioned electron micrograph of strain SC-1 showed that the cell wall was relatively thin and loose causing its gram-negativity and flexible (curved) rod shape. A cell wall analysis showed that strain SC-1 contained menaquinone and branched chain fatty acid, which are widespread in Gram-positive bacteria. As such, these results support that strain SC-1 is close to the Gram-positive bacteria, even though it stained Gram-negatively. The strain SC-1 is a microaerophilic thermophile of the domain Bacteria. Therefore, On the basis of the physiological and molecular properties of the SC-1, it can be named as *Symbiobacterium toebii*.

## The genome sequence of *S. toebii*

The genome sequence of *S. toebii* was determined using a random shotgun library method. The first draft of genome sequence of *S. toebii* was obtained from analysis of 32,474 clones and its sequence redundancy was 5 times. The genome of *S. toebii* constituted a circular chromosome of 3,280,275 base pairs and didn't have a extra chromosomal element (ECE). It contained about 4,107 predicted coding sequence. According to the investigation on a non-redundant protein database (nr_blastx) in NCBI, of these protein coding genes, about 45.6% represented to be encoding well-known proteins and annotated the functional assignment of 1,874 open reading frames (ORFs) and the rest encoded proteins that have unknown function (Table 1). Moreover, the protein sequence of each predicted ORFs from *S. toebii* very resembled each protein sequence of *Bacillus subtilis*. Approximately 66.2% of annotated genes have homology with *Bacillus subtilis*.

Table. 1. Features of *Symbiobacterium toebii* Genome

| | |
|---|---|
| **Genome Sequence Information** | |
| Size | 3,280,275 base pairs |
| Redundancy | 5X |
| Total number of reads | 32,474 clones |
| | |
| **ORF Information** | |
| Predicted coding sequence (containing frame-shift) | 4107 |
| Nr_GeneBank hit (e-value < 1.00E-20) | 1874 |
| Bacillus subtilis hit (e-value < 1.00E-20) | 1241 |
| No database hit | 2233 |

## Industrial Applications of Genomic Resources of *S. toebii*

Genome analysis revealed that *S. toebii* had almost genes encoding enzymes that could be found in other microorganisms except the genes related to branched amino acid-tRNAs synthetases. The structural genes

encoding 25 thermostable enzyme families were first searched in the genomic library of *S. toebii* for industrial use of genome information of *S. toebii* (Table. 2). These enzymes contained 9 different peptidases, 6 esterases, 5 phophatases, 5 oxidases and 3 catalases, and so on. Especially, although *S. toebii* didn't show any lipase activity in the culture broth and cell-free extract, it had two silent genes encoding lipases. This fact indicated that industrially valuable thermostable enzymes could be obtained from genomic library of the *S. toebii* whether the structural genes was expressed or not. Among these useful thermostable enzymes, the genes encoding a thermostable tyrosine phenol-lyase (TTPL) and a

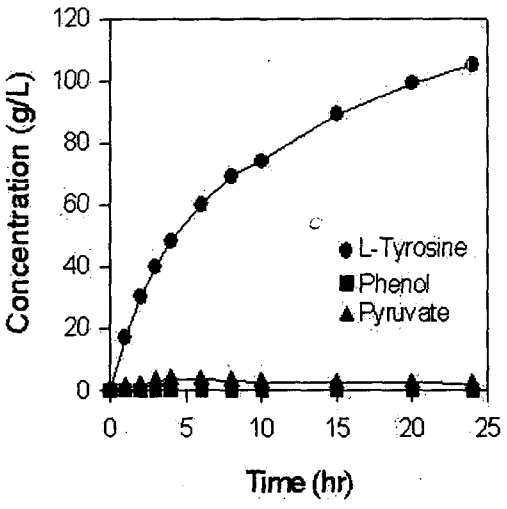


Fig. 4. Time-course of the production of L-tyrosine from phenol and pyruvate with the TTPL

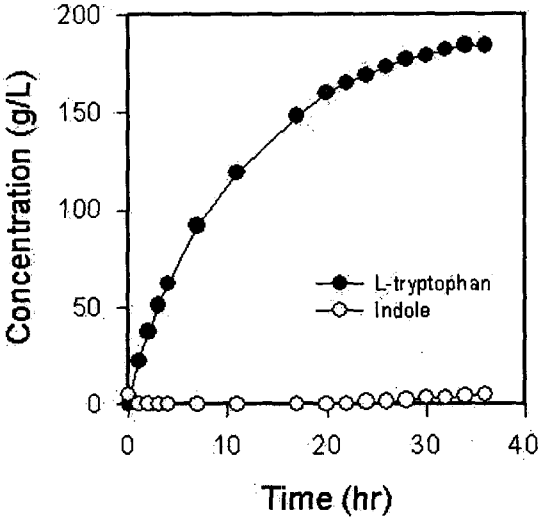


Fig. 5. Time-course of the production of L-tryptophan from indole and pyruvate with the TTNA

thermostable tryptophan indole-lyase (TTNA) were cloned from the genomic DNA of the *S. toebii* by PCR and overexpressed in the recombinant *E. coli* BL21. These two enzymes were used as biocatalysts for the production of L-tyrosine and L-tryptophan. First, the production of L-tyrosine, a medicine for Bacedowi disease, was carried out in 30L of a water-jacketed batch reactor containing initially 10 L of reaction mixture with the TTPL, which catalyzes the cleavage of tyrosine to phenol, pyruvate, and ammonia. The production of L-tyrosine was started with continuous feeding 2 M of phenol and 2M of pyruvate as substrates into 10 L of TPL buffer at 37 °C. The feeding rate of substrates was 20 ml/hr and the composition of the TPL buffer was followed; 77 g of ammonium acetate, 25 mg of PLP, 0.12 g of DTT, 1 g of Na-sulfite, 0.4 g of EDTA, 100 ml of methanol, and 2,000 unit of TTPL per liter. Figure 4 showed the time-curse of the L-tyrosine production from phenol and pyruvate with TTPL. As shown in figure 4, the amount of L-tyrosine produced increased with time, but the concentration of phenol and pyruvate did not change at a low level. Through the overall reaction, 105 g/L of L-tyrosine was produced with productivity of approximately 4.4 g/L/hr.

Second, the TTNA, which catalyzes in vivo degradation of L-tryptophan to indole, pyruvate, and ammonia, was used as a biocatalyst for producing L-tryptophan with a reverse reaction of TTNA. L-

tryptophan used for treatment of mental disorder like depression, anxiety, and premenstrual syndrome, etc Optimal conditions for the production of L-tryptophan were 37 °C, pH 8.5. Figure 5 showed the time-course of the enzymatic production of L-tryptophan. After 24 hr of reaction in above optimal condition, 188 g/L of L-tryptophan (productivity = about 8 g/L/hr) was synthesized from indole, pyruvate, and ammonium chloride.

## Acknowledgements

This work was supported by National Research Laboratory (NRL) program grant NLM0020012 from the Ministry of Science and Technology of Korea.

## References

Carol, J. B., O. White., J. C. Venter & 37 authors (1996) Complete genome sequence of the Methanogenic Archaeon, *Methanococcus jannaschii*. Science 273:1058-1073

DOE Microbial genome program report (2000) http://www.ornl.gov/hgmis/publicat/microbial

Hong S-P, Park J-H, Kim Y-S, Hwang H-J, Rhee S-K, Lee S-G, Sung M-H, Esaki N. (2000) Analysis of the genome of *Symbiobacterium toebii* by pulsed-filed gel electrophoresis. J. Microbiol Biotechnol 10:405-409

Lee S-G, Hong S-P, Choi Y-H, Chung Y-J, Sung M-H (1997) Thermostable tyrosine phenol-lyase of *Symbiobacterium* sp. SC-1: Gene cloning, sequence determination, and overproduction in *Escherichia coli*. Protein Expression purif 11:263-270

Lee S-G, Ro H-S, Hong S-P, Kim E-H, Sung M-H (1996) Production of L-DOPA by thermostable tyrosine phenol-lyase of a thermophilic *Thermosyntrophobacterium* species overexpressed in Recombinant *Escherichia coli*. J. Microbiol Biotechnol 6:98-102

Ohno M, Okano I, Watsuji T, Kakinuma T, Ueda K, Beppu T (1999) Establishing the independent culture of a strictly symbiotic bacterium *Symbiobacterium thermophilum* from its supporting Bacillus strain. Biosci Biotechnol Biochem 63:1083-1090

Ohno M, Shiratori H, Park M-J, Saitoh, Y, Kumon Y, Yamashita N, Hirata A, Nishida H, Ueda K, Beppu T (2000) *Symbiobacterium thermophilum* gen. nov., sp. nov., a symbiotic thermophile that depends on co-culture with a *Bacillus* strain for growth. Int J Syst Evol Microbiol 50:1829-1832

Rhee S-K, Lee S-G, Hong S-P, Choi Y-H, Park J-H, Kim C-J, Sung M-H (2000) Obligately commensal interaction between a new gram-negative thermophile strain SC-1 and *Bacillus* sp. strain SK-1. Extremophiles 4:131-136

Rhee, S.-K., Jeon, C. O., Bae J.-W. & 9 other authors (2001). Characterization of *Symbiobacterium toebii*, an obligate commensal thermophile isolated from compost. *Extremophiles* (In Press)

Suzuki S, Horinouchi S, Beppu T (1988) Growth of a tryptophanase-producing thermophile, *Symbiobacterium thermophilum* gen. nov., sp. nov., is dependent on co-culture with a *Bacillus* sp. J Gen Microbiol 134:2353-2362