Dna Research, 2017, 24(2), 169–177
doi: 10.1093/dnares/dsw055
Advance Access Publication Date: 8 January 2017
Full Paper

OXFORD

Full Paper

Complete genome sequence and expression profile of the commercial lytic enzyme producer *Lysobacter enzymogenes* M497-1

Hideto Takami[1,*], Atsushi Toyoda[2], Ikuo Uchiyama[3], Takehiko Itoh[4], Yoshihiro Takaki[1], Wataru Arai[1], Shinro Nishi[1], Mikihiko Kawai[1,†], Kazuo Shin-ya[5] and Haruo Ikeda[6]

[1]Microbial Genome Research Group, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokohama, Kanagawa, Japan, [2]Comparative Genomics Laboratory, National Institute of Genetics, Mishima, Japan, [3]Laboratory of Genome Informatics, National Institute for Basic Biology, National Institutes of Natural Sciences, Aichi, Japan, [4]Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Meguro-ku, Tokyo, Japan, [5]National Institute of Advanced Industrial Science and Technology, Koto-ku, Tokyo, Japan, and [6]Kitasato Institute for Life Sciences, Kitasato University, Sagamihara, Kanagawa, Japan

[†]Present address: Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Meguro-ku, Tokyo, Japan
*To whom correspondence should be addressed. Tel: +81-45-778-5459. Fax: +81-45-778-5588. E-mail: takamih@jamstec.go.jp

Edited by Dr. Katsumi Isono

Received 14 June 2016; Editorial decision 17 November 2016; Accepted 19 November 2016

Abstract

Lysobacter enzymogenes M497-1 is a producer of commercialized achromopeptidase and is expected to harbour genes encoding various other antimicrobial enzymes. Here, we present the complete sequence of the genome of M497-1 and the expression profiles of the genes for various antimicrobial enzymes. Of the 117 peptidase-encoding genes found in the 6.1-Mb genome of M497-1, 15 genes (aside from the gene encoding the achromopeptidase) were expressed at a level higher than that of the average ribosomal protein genes in the 24-h culture. Thus, the strain was found more valuable than hitherto considered. In addition, M497-1 harbours 98 genes involved in the biosynthesis of various natural products, 16 of which are M497-1-specific across 4 *Lysobacter* species. A gene cluster starting at LEN_2603 through LEN_2673 among the 98 genes closely resembled the lysobactin biosynthesis gene cluster of *Lysobacter* sp. ATCC 53042. It is likely that M497-1 may produce lysobactin or related antibacterial compounds. Furthermore, comparative genomic analysis of M497-1 and four other *Lysobacter* species revealed that their core genome structure comprises 3,737 orthologous groups. Our findings are expected to advance further biotechnological application of *Lysobacter* spp. as a promising source of natural bioactive compounds.

Key words: *Lysobacter enzymogenes*, genome sequence, lytic enzyme, RNA sequencing, metabolic potential

1. Introduction

Lysobacter enzymogenes M497-1 (formerly *Achromobacter lyticus*) was originally isolated from a Japanese soil sample and identified as *A. lyticus* within the class *Betaproteobacteria* by Isono *et al.*[1] It was registered as a producer of the bacteriolytic enzyme lysyl endopeptidase (US patent no. 3,649,454) in 1972.[1] The achromopeptidase (lysyl endopeptidase) produced by this strain2–6 has broad bacteriolytic specificity and has been used for the lysis of anaerobic gram-positive rods, including various *Actinomycetales* and *Bifidobacterium* species,[7] and gram-positive cocci, such as *Micrococcus radiodurans*,[8], *Staphylococcus aureus*,[9] and *Peptococcus saccharolyticus*.[10] The commercialized achromopeptidase is widely employed in metagenomic analyses of the human gut microbiome, which contains anaerobic cocci and rods insensitive to lysozyme, because it enables stable and highly efficient lysis (>90%) of human intestinal microorganisms.[11] Other types of proteolytic enzymes produced by strain M497-1 include α- and β-lytic proteases.[12–14]

Lysobacter species share some properties, including the ability to lyse prokaryotic and eukaryotic microbes, high genomic G + C content (65–70%), and gliding motility, with plant-associated myxobacteria, although *Lysobacter* species have never been observed to form fruiting bodies, which are typical for myxobacteria. Consequently, *Lysobacter* species have been occasionally misidentified as myxobacteria in the older literature,[15–20] but this taxonomic problem was resolved in 1978 by creating a new genus, *Lysobacter*, within the class *Gammaproteobacteria*.[21,22] To clarify the taxonomic position of *A. lyticus* M497-1, which produces lytic enzymes similar to those of *Lysobacter* species,[23–25] we conducted a sequence homology search based on the 16S rRNA gene amplified by PCR from M497-1 DNA remaining in the commercial crude enzyme preparation, and we found 99.9% identity with the 16S rRNA of *L. enzymogenes* (formerly *Myxobacter*) strain 495.[15]

L. enzymogenes is a versatile bacterium producing various types of lytic enzymes specific not only for bacteria[7–10,12–14] but also for eukaryotic cells. Consequently, several *L. enzymogenes* strains have been characterized as biological control agents against plant diseases caused by fungi such as *Bipolaris sorokiniana*[19,20] and *Uromyces appendiculatus*,[26] filamentous protists such as *Pythium aphanidermatum* (class *Oomycota*),[27] and roundworms[28] such as the sugar-beet cyst nematode *Heterodera schachtii* and the root-knot nematode *Meloidogyne javanica*. *L. enzymogenes*-derived chitinases can inhibit conidial germination of plant pathogens and degrade egg shells and hatched juveniles of nematodes[20,28–31], while β-1,3-glucanases from the bacterium can lyse the cell walls of pathogenic fungi.[32,33] Furthermore, *Lysobacter* species are known to synthesize other antimicrobial agents such as tripropeptins[34,35] and lycosin E,[36] which are active against methicillin- and vancomycin-resistant *S. aureus*, respectively.

Over 40 species within the genus *Lysobacter* have been registered in the NCBI taxonomy database, and recently, the complete genomes of 4 species (*L. antibioticus*, *L. capsici*, *L. enzymogenes*, and *L. gummosus*) have been sequenced.[37] The genome sequence data are very useful to elucidate the intragenus as well as intraspecies genomic diversity. Here, we compared the genomes of M497-1 and 4 other *Lysobacter* species to determine the core genomic structure of the genus *Lysobacter*, and explored the genetic, metabolic, and physiological potential of M497-1, which could be exploited in biotechnological applications.

2. Materials and methods

2.1. Strain and cultivation

L. enzymogenes strain M497-1 was kindly supplied by Wako Pure Chemical Industries (Osaka, Japan), Ltd. This strain was deposited to the American Type Culture Collection as *Lysobacter* sp. (ATCC 21456). M497-1 was aerobically grown at 30 °C in 500-ml flasks containing 100 ml of medium for 24 or 48 h. For DNA analysis, the bacteria were grown in CY medium containing 0.5% casitone and 0.1% yeast extract, and for RNA expression analysis, they were cultured in medium containing 1.5% glycerol, 0.3% NaCl, 0.5%, L-glutamate monohydrate (pH 7.4), and 1.5% cotton seed meal (Traders Protein, Southern Cotton Oil Company, Memphis, TN, USA).

2.2. DNA and RNA purification

Genomic DNA was extracted and purified using the NucleoSpin Tissue kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) according to the manufacturer's protocol. RNA was extracted and purified from 24-h and 48-h M497-1 cultures using the RNeasy Protect Bacteria Mini kit and QIAzol Lysis Reagent (Qiagen, Tokyo, Japan) according to the manufacturer's protocol. DNase I-treated RNA was re-purified using the RNeasy Mini Cleanup Kit (Qiagen) and rRNA was removed using the Ribo-Zero rRNA Removal Kit (Gram-Negative Bacteria) (Epicentre, Madison, WI, USA).

2.3. Sequencing, assembly, and data analyses

For whole-genome shotgun sequencing, we combined the Roche GS FLX (Roche Diagnostics K.K., Tokyo, Japan) and Sanger sequencing technologies, which produced ~53- and 23-fold genome coverage, respectively. First, GS FLX reads with an average length of 231 bp were assembled using the Newbler programme version 2.3, and 385 contigs with an N50 length of 33,738 bp were generated. Next, we performed end-sequencing of 77,952 and 1,920 clones from 2–3 to 35–40 kb (fosmid) insert-size libraries, respectively, using an ABI 3730xl sequencer (Thermo Fisher Scientific K.K., Yokohama, Japan), and hybrid assembly was carried out by the KB Basecaller and Phrap/Consed programmes.[38] The remaining gaps between contigs were filled in using PCR and Sanger sequencing of amplified products. In addition, low-quality genomic regions were sequenced to increase the genome coverage.

RNA-seq libraries were prepared from 24- to 48-h cultures according to the standard Illumina protocol and cDNA libraries were checked for quality and quantity using the DNA-100 kit (Agilent Technologies Japan, Ltd., Tokyo, Japan) and a 2100 Bioanalyzer. Each library was sequenced with the Illumina Sequencing Kit v2 on one lane of a MiSeq desktop sequencer (Illumina K.K., Tokyo, Japan) to obtain 150-bp average paired-end reads. The value of Reads Per Kilobase per Million mapped reads (RPKM) was calculated according to the standard method. The RPKM ratio was calculated by dividing the RPKM of each gene by the mean RPKM of all ribosomal proteins, and was used to determine relative gene expression levels. Promoter-like sequences were detected with GENETYX-MAC version 18 (Tokyo, Japan) and PePPER[39] (http://pepper.molgenrug.nl/index.php/prokaryote-promoters). Peptidase genes were identified by BLAST searches and from the MEROPS database of proteolytic enzymes; MEROPS identifiers were assigned on the basis of a phylogenetic tree: an identifier was assigned to all sequences derived from the same node as a holotype.[40]

2.4. Orthologous analysis and estimation of gene clusters for biosynthesis of natural product

Orthologous groups (OGs) for M497-1 and 4 other *Lysobacter* species[37] were constructed using the rapid classification programme DomClust[41] on the Microbial Genome Database (MBGD) server.[42] A core genome is defined as a set of genes (OGs) syntenically conserved in at least half of the compared strains. In our study, a set of

genes in the syntenic regions shared by at least three strains was defined as the core genome for the five analysed *Lysobacter* strains. A set of syntenic regions and the consensus order of OGs in these regions designated 'core genome structure' were created using the CoreAligner programme[43] based on conserved linkages between orthologous genes in each chromosome. Among the OGs in the core genome, those present in all the genomes were called 'universal core genome'. Non-ribosomal protein synthase (NRPS)- and polyketide synthase (PKS)-encoding genes were manually annotated using web-based prediction programmes.[44–47] Gene clusters related to biosynthesis of bioactive natural product were estimated based on gene direction and continuity in the M497-1 genome and compared with previously identified gene clusters.

2.5. Evaluation of the metabolic and physiological potential

The metabolic and physiological potential of *L. enzymogenes* M497-1 was investigated using Metabolic and Physiological Potential Evaluator (MAPLE) ver. 2.1.[48] Genes were mapped to functional modules defined by the Kyoto Encyclopedia of Genes and Genomes (KEGG) (pathways: 255, complexes: 284, functional sets: 149, signatures: 40) and the module completion ratio (MCR) was calculated according to a Boolean algebra-like equation described previously.[49] The MCR patterns of M497-1 were compared with those of other *Lysobacter* species.

3. Results and discussion

3.1. General features of the *L. enzymogenes* M497-1 genome

The genome of *L. enzymogenes* M497-1 consists of a single circular chromosome (6,096,022 bp) with a mean G + C content of 69.4% (Fig. 1A). We identified 4,891 protein-coding sequences (CDSs), 55 transfer RNA genes, and 6 rRNA genes comprising 2 rRNA operons (Table S1). Although there was no significant difference in genome size between *L. enzymogenes* strains M497-1 and C3, the former has 638 CDSs fewer than the latter. M497-1 and C3 shared 4,043 OGs, while they had 488 and 1,022 unique genes, respectively. Among 485 M497-1 genes unique to the 5 *Lysobacter* spp. (*L. enzymogenes* M497-1 and C3, *L. antibioticus* 76, *L. capsici* 55, and *L. gummosus* 3.2.11), 8 encoded peptidases and 16 encoded NRPSs and PKSs, while the function of 338 unique genes (70%) is not known. Among the functionally unknown genes, 11 showed an RPKM ratio higher than 2 in 24-h culture; LEN_1053 had the highest RPKM ratio (186) among all genes identified in the M497-1 genome. The original RPKM values and RPKM ratios for all genes are listed in Supplementary Table S2.

The core genome structure of the five *Lysobacter* strains was defined using CoreAligner[43]; it contains a set of well-conserved syntenic OGs (Supplementary Fig. S1). Out of 4,891 CDSs in the M497-1 genome, 3,593 were included in the core genome comprising 3,737 OGs; among them, 3,081 were shared by all 5 strains and thus, represented universal core gene clusters. In total, 3,046 OGs show one-to-one correspondence (Fig. 1B). Types III and IV secretion pathways are highly conserved in the M497-1 genome as well as in those of other *Lysobacter* species (Supplementary Fig. S2).[37] All genes of *L. enzymogenes* M497-1 related to type III secretion (LEN_4800, LEN_4801, LEN_4803, and LEN_4806–4815) were included in the universal core gene cluster (Table S2). However, three genes related to the type IV secretion pathway (LEN_2216–2218) were not included in the universal core as they were not identified as orthologs in the *L. antibioticus* 76 genome by the MBGD server, although the gene organization of the adjacent region was well conserved. The RPKM ratios of genes related to type IV secretion in the M497 genome varied; 6 out of 10 genes (LEN_2207, LEN_2209, LEN_2211, and LEN_2213–2215) demonstrated almost equal or higher expression than the average level for ribosomal proteins in 24-h culture. In particular, LEN_2213 and LEN_2214 showed RPKM ratios of ~9–10; however, the expression level dramatically decreased in 48-h culture, indicating that type IV secretion is active in the exponential growth phase. In contrast to the type IV secretion genes, most genes related to the type III secretion pathway showed 5- to 10-fold lower expression levels than ribosomal proteins.

Iron uptake is vital for almost all bacteria because iron is used as a cofactor in key metabolic processes, including deoxyribonucleotide synthesis, oxidative phosphorylation, and electron transport. To survive under iron-limiting conditions, bacteria have developed sophisticated iron acquisition strategies based on the secretion of iron-chelating siderophores.[50] TonB-dependent transporters are bacterial outer membrane receptor proteins that bind and translocate siderophores as well as vitamin B_{12}, nickel complexes, and carbohydrates.[51] We found that the *Lysobacter* strains possess a remarkable number of genes encoding the TonB-dependent family of proteins, of which 52 and 41 were included in the core and universal core gene clusters, respectively (Supplementary Table S2). Considering that *Lysobacter* spp. are inhabitants of soil and water, which are characterized by low bioavailability of iron, the high number of TonB-dependent proteins should be helpful for iron acquisition. In fact, seven genes included in the core OGs (LEN_0009, LEN_0351, LEN_3594, LEN_3807, LEN_4109, LEN_4624, and LEN_4786) in the M497-1 genome showed an expression level equal to or higher than that of ribosomal proteins in 24-h culture; among them, the highest RPKM ratio (more than 10) was observed for LEN_4109.

3.2. Bacteriolytic enzymes

One hundred seventeen genes encoding peptidases, including the commercialized lysyl endopeptidase (LEN_1308), were identified in the M497-1 genome. These peptidase genes were classified into 56 families according to the MEROPS peptidase database.[40] The largest family, S08A, comprised 21 peptidases, and the S01A family contained LEN_1308 together with 2 other peptidases, LEN_1307 and LEN_1765 (Supplementary Table S4). Among all peptidase genes, 80 were included in the universal core gene clusters (Fig. 2A and Supplementary Table S3) and 16 showed an expression level equal to or higher than the average level of ribosomal proteins in 24-h culture (Supplementary Fig. S3); LEN_4116 (β-lytic protease) of the M23A family showed an RPKM ratio of ~30. However, the expression level of these 16 genes dramatically decreased after 48 h (Supplementary Fig. S3). In gram-positive bacilli, among which are various industrial enzyme producers, protease production usually starts from the early stationary phase of cultivation (20–24 h), and the productivity is maintained even at the late stationary phase (48 h).[52] In contrast to bacilli, *L. enzymogenes* M497-1 easily lyse after reaching the stationary phase; cell lysis was actually observed after 24 h of cultivation (data not shown), likely due to autolysis by autoproduced peptidases.

LEN_1308, located in a peptidase island comprising nine continuous peptidase genes, showed RPKM ratios of 14 and 1.6 in the 24- and 48-h cultures, respectively, and the expression of the neighbouring LEN_1309 and LEN_1310 genes also exceeded the average level of ribosomal proteins (Fig. 2C); moreover, potential σ^{54}-dependent

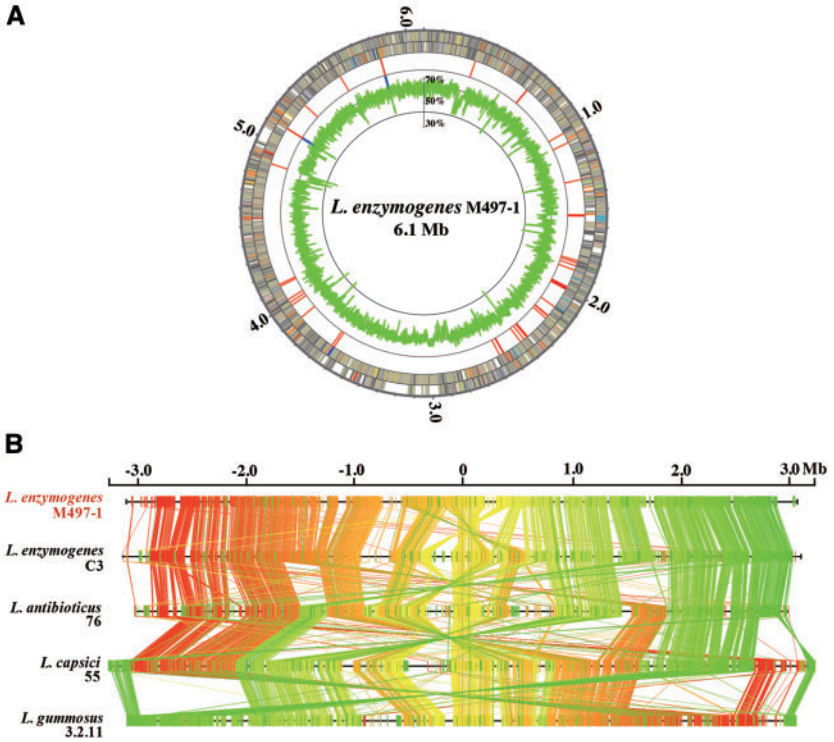


Figure 1. Genomic structure of *Lysobacter* sp. **(A)** Genome map of *L. enzymogenes* M497-1. The distribution of CDSs is indicated by coloured boxes according to the KEGG functional modules and transcription direction; the first two outer circles represent sense and antisense strands, respectively, and the third and fourth circles indicate the distribution of tRNAs and rRNAs, respectively. The innermost circle (in green) represents variations in the G + C content (%) of the genome. Red, energy metabolism (pathway); pink, carbohydrate and lipid metabolism (pathway); orange, nucleotide and amino acid metabolism (pathway); magenta, secondary metabolism (pathway); blue, energy metabolism (complex); magenta and cyan, genetic information processing; soft blue, environmental information processing (complex); yellow green, metabolism (functional set); light green, environmental information processing (functional set); dark green, cellular processes (functional set); yellow, gene set (signature); light gray, hypothetical conserved protein; dark gray, hypothetical protein; tan, KEGG Orthology not assigned to modules. **(B)** Genes in the universal core genome structure for five *Lysobacter* species constructed on the basis of the consensus arrangement of conserved orthologs. Gene location in each chromosome is indicated; ortholog groups are denoted by coloured lines; only universally conserved ortholog groups with one-to-one correspondence are shown. To visualize chromosomal rearrangement, a colour gradient scale (red to yellow to green) is used according to the location on the M497-1 chromosome. Origins of replication are at the center.

promoter sequences were identified upstream of these genes (Supplementary Fig. S3). Among the other genes in the peptidase island (LEN_1301–1306), another σ^{54}-dependent promoter-like region was detected for LEN_1301, suggesting polycistronic expression of the LEN_1301–1306 region; however, expression was very low compared with the average level of ribosomal proteins (Fig. 2C). On the other hand, expression of 4 peptidase genes (LEN_1590, LEN_1765, LEN_1799, and LEN_3019) was detected, while no putative promoter-like sequences were identified (Supplementary Fig. S3). Our manual search for other types of promoter-like regions was not successful either, probably because promoters depending on different σ factors are likely to have significant sequence variability.

3.3. Lytic enzymes having activity against eukaryotic cells

L. enzymogenes is known to protect plant from diseases caused by various fungi, filamentous protists, and nematodes. Chitinase is a hydrolytic enzyme cleaving glycosidic bonds in chitin, a component of fungal cell walls and of exoskeleton of lower animals, including nematodes. *L. enzymogenes* M497-1 has 3 chitinase-encoding genes belonging to glycoside hydrolase families 18 and 19; all of these genes were included in the universal core OGs (Fig. 3A). Among them, LEN_2961 demonstrated significant expression in 24-h culture, as evidenced by its RPKM ratio of 3.9, while the other two genes (LEN_3788 and LEN_4456) were expressed at levels below the average of ribosomal proteins; upstream of each gene, putative σ^{54}-dependent promoter-like sequences were detected (Fig. 3B).

The M497-1 genome also contains 4 endoglucanase/β-1,3-glucanase-encoding genes belonging to glycoside hydrolase families 6, 8, 16 and 64; 2 of these (LEN_0845 and LEN_3546) were included in the universal core OGs (Fig. 3A). *L. enzymogenes* C3 lacks an orthologous gene corresponding to LEN_1176 (family 6), while *L. antibioticus* 76 and *L gummosus* 3.2.11 do not have orthologs of LEN_4406 (family 16) (Fig. 3B). Except for LEN_3546, endoglucanase/β-1,3-glucanase-encoding genes were significantly expressed (relative RPKM ratio of 1.5 and higher in 24-h culture); the highest levels were observed for LEN_0845 (family 8), which demonstrated RPKM ratios of almost 10 and 2 in 24-h and 48-h cultures, respectively. However, no σ^{54}-dependent promoter-like sequences could be detected for these genes.

3.4. Natural products

Although natural products from gram-positive actinomycetes have been extensively studied,[53] those from gram-negative bacteria have received little attention. However, a number of *Lysobacter* spp. are currently recognized as promising antibiotic producers.[54] In the *L.*

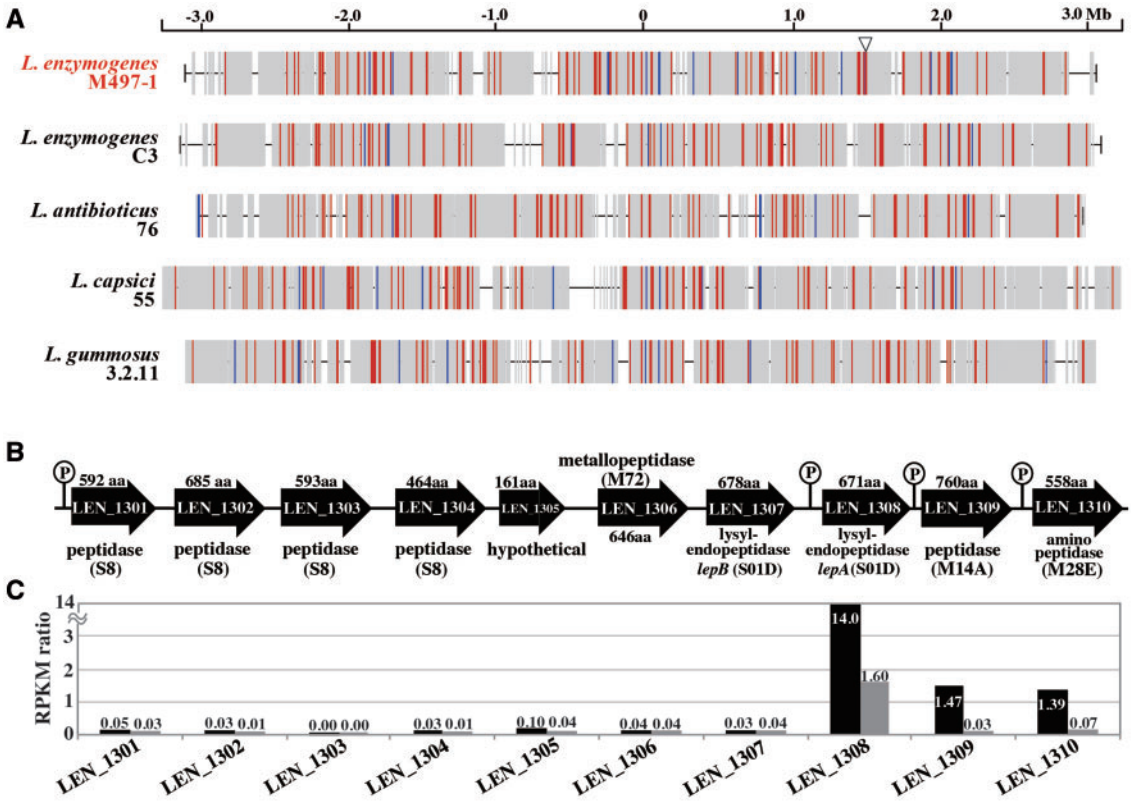
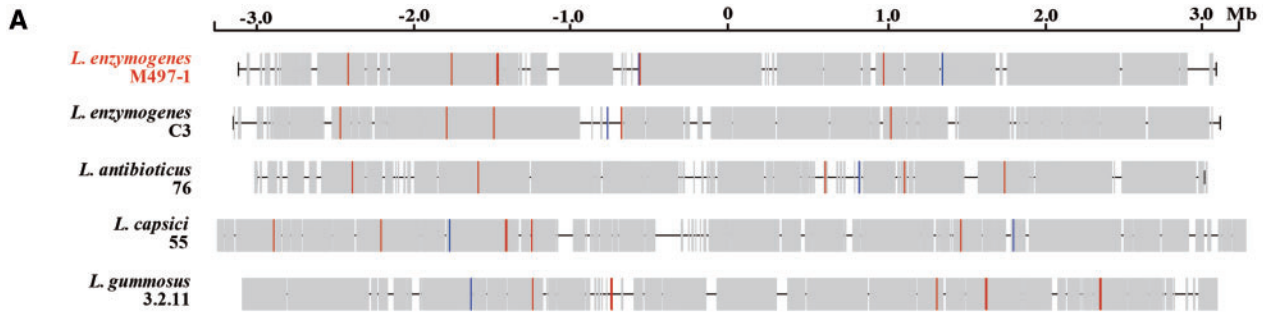


Figure 2. Distribution and expression of peptidase genes in *Lysobacter* strains. **(A)** Distribution of peptidase genes in the genomes of five *Lysobacter* strains. The backbone of each genome is composed of 3,081 universal core OGs, which are represented as gray lines. Red and blue lines represent peptidase genes included and not included in the universal core genome structure, respectively. The triangle shows the position of peptidase islands in the M497-1 genome, detailed gene organization of which is shown in B. **(B)** Gene organization in the peptidase island of the M497-1 genome. Characters in parentheses indicate family names assigned by MEROPS[40], and P indicates the position of promoter-like sequences. **(C)** Expression patterns (RPKM ratios) of the genes in the peptidase island in 24-h (black) and 48-h (grey) M497-1 cultures (see also Supplementary Table S2).

enzymogenes M497-1 genome, we found 98 genes related to the biosynthesis of natural products such as lantibiotics, non-ribosomal peptide and polyketide compounds, and siderophores (Supplementary Table S5); to the best of our knowledge, this is the highest number of such genes identified in *Lysobacter* genomes. However, none of these genes, except for LEN_4535, was included in the universal core OGs, although some of them had orthologous relationships to those of other *Lysobacter* strains. Among the 98 genes, 31 encoded NRPSs, type I PKSs, and NRPS/PKS hybrids comprising various catalytic domains, and 12 of them, LEN_2075-2077, LEN_2515–2517 and LEN_2548–2551 (NRPS/PKS), LEN_2803 (type I PKS), and LEN_4535 (NRPS) were well conserved at similar positions in the genomes of *L. enzymogenes* stains M497-1 and C3 (Fig. 4 and Supplementary Table S6). On the other hand, M497-1 was found to have many unique genes, especially for type I PKS and NRPS/PKS hybrids, which have not yet been identified in other fully sequenced *Lysobacter* spp. (Supplementary Table S6). Furthermore, 9 out of 20 genes responsible for NRPS were also unique in M497-1. These unique genes might have been acquired through horizontal gene transfer because they showed 40–60% sequence identity with those of several species within phylogenetically distant genera *Rhodococcus*, *Polyangium, Bradyrhizobium*, except for *Xanthomonas*, within the same family. Unfortunately, significant expression was observed only for the NRPS cluster comprising LEN_4535 and LEN_4536, which showed RPKM ratios of 4–6 in 24-h culture and 2.5 in 48-h culture although the product was not identified in the culture medium. Although genes encoding types II and III PKSs (which are involved in the biosynthesis of aromatic compounds) and genes involved in the biosynthesis of terpene compounds have been frequently found in *Actinomycetales* genomes, such genes were not identified in the M497-1 genome. Interestingly, the region LEN_2603–LEN_2673 (ca. 145 kb) was extremely similar to that containing a gene cluster for lysobactin biosynthesis of *Lysobacter* sp. ATCC 53042, which is a patent strain (accession no. JF412274). The lysobactin biosynthesis cluster comprising LG3211_2475, 2476 is also present in the *L. gummosus* 3.2.11 genome,[37] and these genes showed significant identities (>70%) to LEN_2615 and LEN_2613. Through the analysis of a series of genes involved in lysobactin biosynthesis in *Lysobacter* sp. ATCC 53042, it was suggested that the region LEN_2608–LEN_2615 is responsible for the biosynthesis of lysobactin or related compounds.[55] Although 2 NRPSs (LybA and LybB) of *Lysobacter* sp. ATCC 53042 are involved in the formation of the peptide backbone of lysobactin, 3 NRPSs (LEN_2613–2615) identified in the *L. enzymogenes* M497-1 genome were found to correspond to LybA and LybB; LEN_2615 corresponds to LybA, while LybB is divided into two NRPSs, LEN_2614 and LEN_2613. Thus, *L. enzymogenes* M497-1 is presumably able to produce lysobactin or related antibacterials. Notably, gene clusters encoding other antibiotics identified in *Lysobacter* species, such as heat-stable antifungal factor (HSAF)/dihydromaltophilin, WAP-8294A2 (anti-MRSA), and phenazine,[37] were not detected in M497-1.



B

enzyme name	catregory	*L. gumosus* 3.2.11	*L. capsici* 55	*L.antibioticus* 76	*L. enzymog.* C3	*L. enzymog.* M497-1	RPKM ratio 0 10	predicted promoter-like sequence
endoglucanase/ β-1,3-glucanase	glycoside hydrolase family 8	LG3211_4319	LC55x_4508	LA76x_0932	GLE_4374	LEN_0845	9.82 / 1.98	n.d
	glycoside hydrolase family 6	LG3211_3970	LC55x_4173	LA76x_0665		LEN_1176	2.37 / 0.12	n.d
	glycoside hydrolase family 16		LC55x_1557		GLE_0490	LEN_4406	1.5 / 0.11	n.d
	glycoside hydrolase family 64	LG3211_1435	LC55x_3780	LA76x_3779	GLE_1403	LEN_3546	0.24 / 0.16	TCTAAGCGGAATGGGCGCTCAGGGTATCAT
chitinase	glycoside hydrolase family 19	LG3211_2114	LC55x_3117	LA76x_3079	GLE_2053	LEN_2961	3.9 / 0.08	ACGAATTCAGCTTCATCGCGCCGCTACATT
	glycoside hydrolase family 18	LG3211_1156	LC55x_1252	LA76x_1507	GLE_1126	LEN_3788	0.83 / 0.08	TCGAAGGAAGGTCGTCGAACCTTCCAAT
	glycoside hydrolase family 18	LG3211_4770	LC55x_4660	LA76x_0490	GLE_0401	LEN_4456	0.52 / 0.02	TTGACAGGCGGGGCGCCGCCCACCAAAGT

Figure 3. *Lysobacter* genes encoding lytic enzymes active against eukaryotic cells. **(A)** Distribution of the genes encoding endoglucanase/*β*-1,3-glucanase and chitinase (genome backbones are the same as in Fig. 2); red and blue lines indicate the lytic enzyme genes included and not included in the universal core genome structure, respectively. **(B)** Orthologous relationships of lytic enzymes among five *Lysobacter* strains and expression patterns of the orthologs identified in M497-1. Red and blue boxes correspond to red and blue lines, respectively, in A; green and yellow bars show RPKM ratios in 24- and 48-h cultures, respectively. Promoter-like sequences detected only by PePPER[39] are shown in blue, and those identified by both PePPER and GENETYX-MAC are shown in black. n.d., not detected.

3.5. Metabolic and physiological potential

To compare the metabolic and physiological potential of M497-1 and other *Lysobacter* strains, the completion ratios of the four types of KEGG functional modules (pathway, structural complex, functional set and signature modules) were calculated with MAPLE (Supplementary Table S7).[48] Most MCR patterns were similar among the 5 *Lysobacter* strains (Supplementary Fig. S4); however, several modules demonstrated species-specific differences (Fig. 5). Among them, there were seven pathway modules related to amino acid metabolism, nitrogen metabolism, carbohydrate metabolism, and terpenoid backbone biosynthesis. In the ornithine biosynthesis module comprising 5 reaction steps, *L. enzymogenes* M497-1 and *L. gummosus* 3.2.11 lacked amino-acid N-acetyltransferase (*argA*) assigned to the first reaction step. The module for dissimilatory nitrate reduction comprising two reaction steps was complete only in *L. antibioticus* 76, while the enzymes for each reaction step, nitrate reductase and nitrite reductase, were absent in the other 4 species. Except for *L. enzymogenes* strains, the other three *Lysobacter* spp. completed the module for D-galactonate degradation, which is one of the rare modules, i.e. completed by less than 10% of prokaryotes.[49]

The differences in MCR patterns among *Lysobacter* strains were also observed in structural-complex and functional-set modules. As a conspicuous example, only *L. antibioticus* 76 completed the modules for RTX (repeats in toxin) transport belonging to the type I secretion system, putrescine transport, phospholipid transport, aminoglycoside-resistant protease HtpX, and the osmotic stress response (EnvZ-OmpR) two-component regulatory system (Fig. 5). In addition, the modules for type VI secretion and simple sugar transport were completed only by *L. enzymogenes* C3 and *L. capsici* 55, respectively. Moreover, strain-specific differences between M497-1 and C3 were detected in 3 rare modules: the α-hemolysin/cyclolysin transport system and two-component regulatory systems associated with central carbon metabolism (BarA-UvrY) and chemosensing (WspE-WspRF). Thus, we could detect exact differences in the metabolic and physiological potential among five closely related *Lysobacter* species.

3.6. Regulatory factors

The production of bioactive products in *Lysobacter* strains is regulated by intracellular signaling mediated by cyclic adenosine monophosphate-receptor-like protein (Clp) and the diffusible signal factor (DSF)-dependent system.[56–58] The Clp-encoding gene was identified in all *Lysobacter* genomes, including M497-1 (LEN_0938), and the deduced gene products were 99–100% identical to each other. LEN_0938 was highly expressed as evidenced by its RPKM ratio of 6.6, and the expression of various Clp-regulated genes, including those encoding lytic enzymes such as chitinase, endoglucanase/*β*-1,3-glucanase, and peptidase, was also markedly induced (Supplementary Table S2). These data are consistent with previous reports that Clp upregulated the expression of chitinase, *β*-1,3-glucanse, peptidase, and antifungal factor in *L. enzymogenes* C3 and OH11.[57,58]

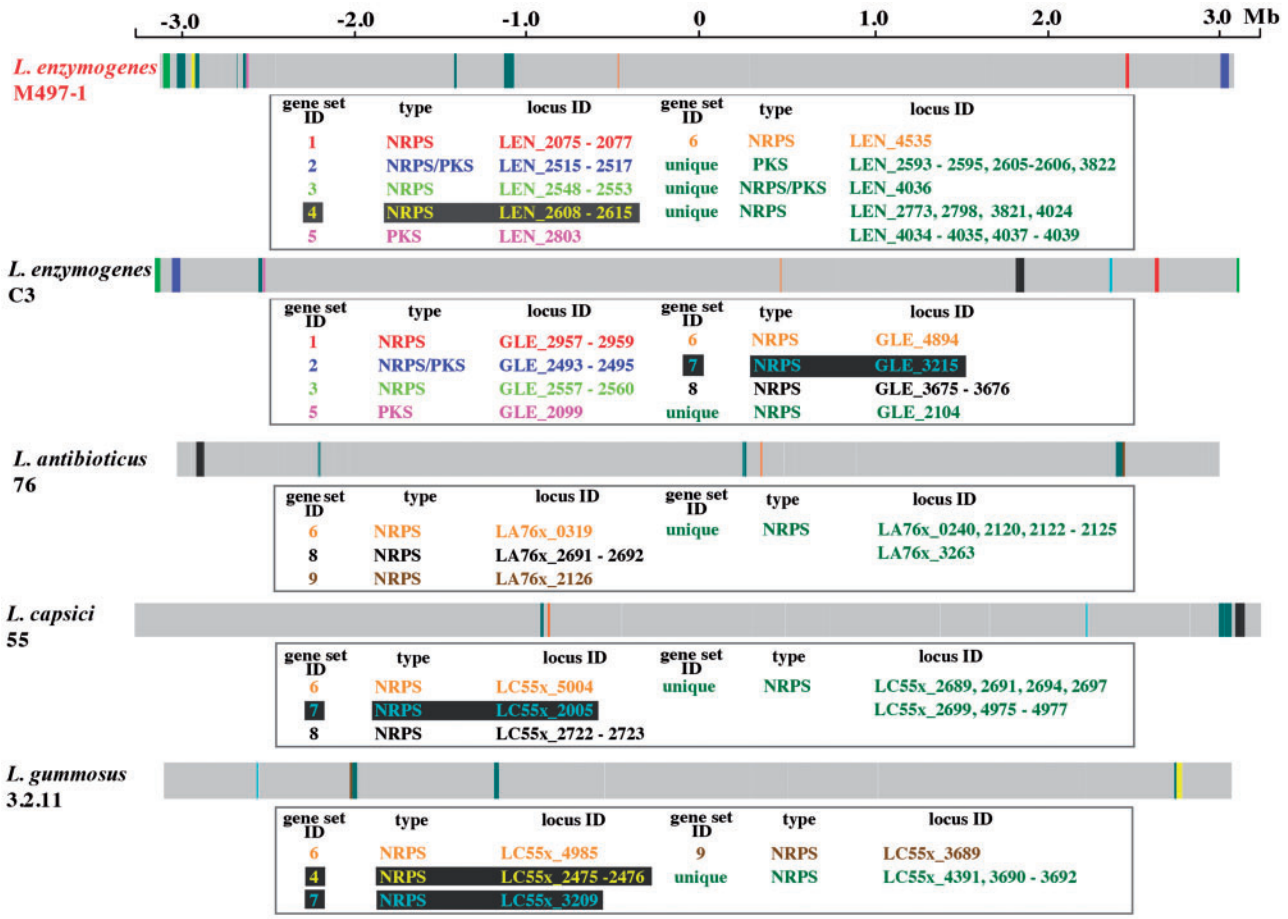


Figure 4. Distribution of genes encoding NRPS and PKS with core catalytic domains for secondary metabolite biosynthesis. Gene sets with the same ID number (1–9) are orthologous; gene sets unique for each strain (no ID number) are indicated. The detailed core domain structure of each gene is shown in Supplementary Table S5.

Genes encoding components of the DSF-dependent system (*rpfG*, *rpfC*, *rpfF*, and *rpfB*), which is widely conserved among *Lysobacter* species (with the exception of *L. capsici*), were also identified in the M497-1 genome (LEN_2756–2759). A previous study did not detect genes corresponding to *rpfC* and *rpfF* in *L. capsici* 55; however, two genes (LC55x_5419 and LC55x_5418) were assigned to the *rpfF* and *rpfC* OGs, respectively, by MBGD analysis.[37] These genes showed 30–40% sequence similarity to phylogenetically distant species within different classes such as *Betaproteobacteria* and *Alphaproteobacteria*, suggesting that they might have been acquired by *L. capsici* 55 through horizontal gene transfer. *L. enzymogenes* OH11 produces a DSF-like protein, LeDSF3, which positively regulates HSAF biosynthesis through NRPS and the RpfC/RpfG-Clp pathway.[58] However, in the M497-1 genome, the genes responsible for the DSF-dependent system (LEN_2756–2759) showed very low expression when compared with Clp (RPKM ratio 0.2–0.6 versus 6.6, respectively), and genes related to HSAF biosynthesis regulated by the DSF-dependent system were not identified.

With regard to antibiotic production pathways, we found 16 genes related to type I PKS and NRPS to be unique to M497-1 (Supplementary Table S6). Additionally, all of them were silent, although three other orthologous genes (LEN_4535, LEN_2614, and LEN_2615) showed an expression level greater than half the average expression level of ribosomal proteins in a 24-h culture (>0.58). Further studies focusing on the regulatory mechanisms underlying antibiotic production in *L. enzymogenes* M497-1 will be required with regard to not only unique genes but also other orthologous genes to identify novel antibiotics synthesized by this strain.

In conclusion, *L. enzymogenes* M497-1 genome sequencing and expression profiling performed in this study revealed 16 unique genes for biosynthesis of natural product and 15 highly expressed genes encoding peptidases except for commercialized achromopeptidase. It also represented the possibility to produce lysobactin or related antimicrobial compounds. Thus, our findings are expected to promote further biotechnological application of *Lysobacter* spp. as a promising source of natural, bioactive compounds.

Acknowledgements

We thank Mr T. Hara of Wako Pure Chemical Industries, Ltd. for his valuable contribution to this work.

Data availability

The nucleotide sequences determined in this study were deposited in DDBJ/EBI/NCBI database. The accession number for *L. enzymogenes* M497-1 whole genome sequence is AP014940 and that for RNA sequences (RNA-seq) of 24- and 48-h cultures is DRA003922.

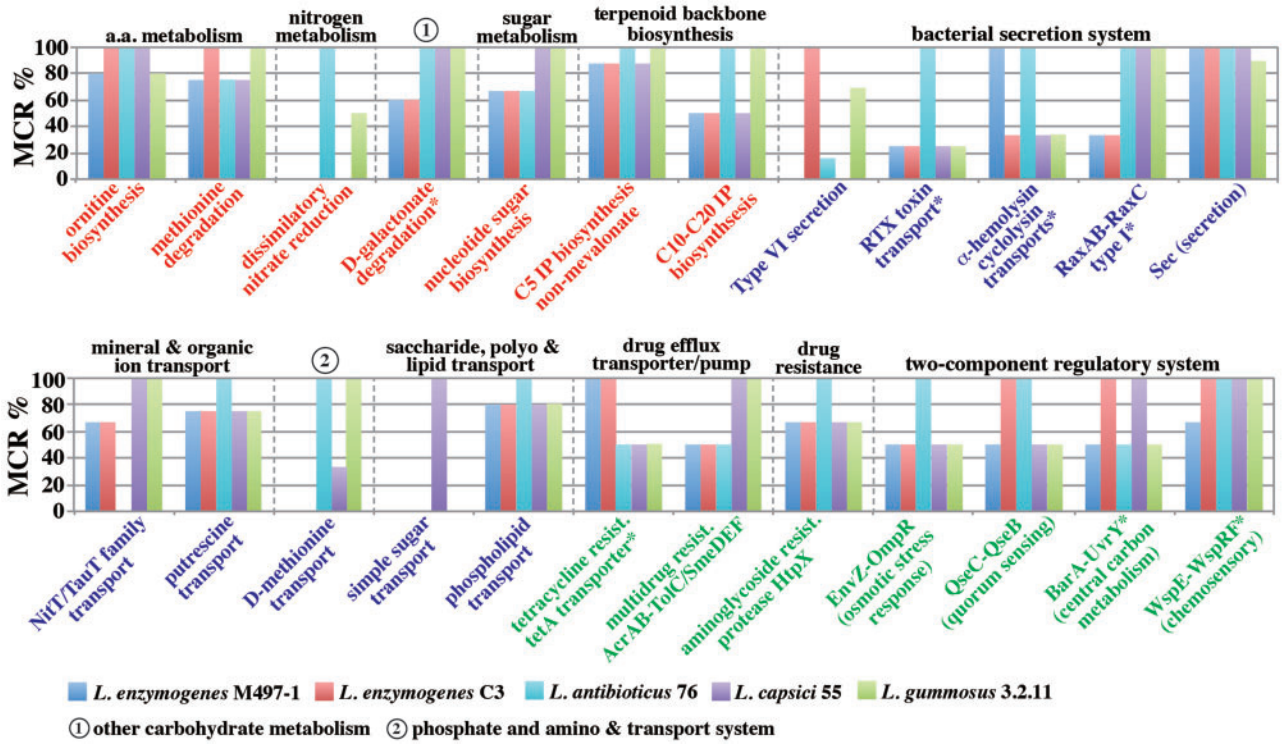


Figure 5. Major differences in metabolic and physiological potential among *Lysobacter* strains. The MCR of the KEGG functional modules was calculated using MAPLE[48]; modules with different MCR patterns among *Lysobacter* strains are highlighted. First 7 modules, pathways; next 10 modules, structural complexes; last 7 modules, functional sets. Modules defined as 'rare' are indicated with an asterisk.[49]

Conflict of interest

None declared.

Accession numbers

AP014940

Supplementary data

Supplementary data are available at www.dnaresearch.oxfordjournals.org.

Funding

This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan to H.T. (No. 20310124 and 26550053) and by a grant from the Cross-ministerial Strategic Innovation Promotion Programme to HT and WA. This work was also supported in part by grants from the collaborative research programme of National Institute for Basic Biology (NIBB) to H.T. and I.U. (No. 13-360, 14-359, and 15-355). Computational resources were provided partly by the Data Integration and Analysis Facility of NIBB.

References

1. Isono, M., Takahashi, T. and Yamazaki, Y. 1972, *Bacteriolytic enzyme and process for the production thereof*. US Patent 3,649,454.
2. Masaki, T., Nakamura, K., Isono, M. and Soejima, M. 1978, A New proteolytic enzyme from *Achromobacter lyticus* M497-1. *Agric. Biol. Chem.*, **42**, 1443–5.
3. Masaki, T., Tanabe, M., Nakamura, K. and Soejima, M. 1981, Studies on a new proteolytic enzyme from *Achromobacter lyticus* M497-1. I. Purification and some enzymatic properties. *Biochim. Biophys. Acta*, **660**, 44–50.
4. Masaki, T., Tanabe, M., Nakamura, K. and Soejima, M. 1981, Studies on a new proteolytic enzyme from *Achromobacter lyticus* M497-1. II. Specificity and inhibition studies of *Achromobacter* protease I. *Biochim. Biophys. Acta*, **660**, 51–5.
5. Masaki, T., Suzuki, H. and Soejima, M. 1986, Purification and some properties of *Achromobacter* protease Ia from *Achromobacter lyticus* M497-1. *Agric. Biol. Chem.* **50**, 3087–91.
6. Ohara, T., Makino, K., Shinagawa, H., Nakata, A, Norioka, S. and Sakiyama, F. 1989, Cloning, nucleotide sequence, and expression of *Achromobacter* protease I gene. *J. Biol. Chem.* **264**, 20625–31.
7. Barsotti, O., Morrier, J. J., Freney, J., et al. 1988, Achromopeptidase for rapid lysis of oral anaerobic Gram-positive rods. *Oral Microbiol. Immunol.*, **3**, 86–8.
8. Nakamura, K., Okazawa, Y., Soejima, M. and Masaki, T. 1973, Lysis of *Micrococcus radiodurans*. *Agric. Biol. Chem.*, **37**, 2667–8.
9. Horinouchi, S., Uozumi, T., Beppu, T. and Arima, K. 1977, A new isolation method of plasmid deoxyribonucleic acid from *Staphylococcus aureus* using a lytic enzyme of *Achromobacter lyticus*. *Agric. Biol. Chem.*, **41**, 2487–9.
10. Ezaki, T. and Suzuki, S. 1982, Achromopeptidase for lysis of anaerobic Gram-positive cocci. *J. Clinic. Microbiol.* **16**, 844–6.
11. Morita, H., Kuwahara, T., Ohshima, K., et al. 2007, An improved DNA Isolation method for metagenomic analysis of the microbial flora of the human intestine. *Microbes Environ.* **22**, 214–22.
12. Li, S., Norioka, S. and Sakiyama, F. 1990, Molecular cloning and nucleotide sequence of the β-lytic protease gene from *Achromobacter lyticus*. *J. Bacteriol.* **172**, 6506–11.
13. Li, S., Norioka, S. and Sakiyama, F. 1997, Purification, Staphylolytic activity, and cleavage sites of α-lytic protease from *Achromobacter lyticus*. *J. Biochem.* **122**, 772–8.
14. Li, S., Norioka, S. and Sakiyama, F. 1998, Bacteriolytic activity and specificity of *Achromobacter* β-lytic protease. *J. Biochem.* **124**, 332–9.
15. Damoglou, A. P., Allen, L. C., Roy, C. and Whitaker, D. R. 1972, The β-lytic protease of *Myxobacter* 495: aspects of its structure. *Proc. Can. Fed. Biol. Soc.* **15**, 27.

16. Ensign, J. C. and Wolfe, R. S. 1965, Lysis of bacterial cell walls by an enzyme isolated from a *Myxobacter*. *J. Bacteriol.* **90**, 395–401.
17. Ensign, J. C. and Wolfe, R. S. 1966, Characterization of a small proteolytic enzyme which lyses bacterial cell walls. *J. Bacteriol.* **91**, 524–34.
18. Whitaker, D. R. 1965. Lytic enzymes of *Sorangium* sp., isolation and properties of the α-lytic and β-lytic proteases. *Can J. Biochem.* **43**, 1935–54.
19. Zhan, Z. and Yuen, G. Y. 1999, Biological control of *Bipolaris sorokiniana* on tall fescue by *Stenotrophomonas maltophila* strain C3. *Phytopathology* **89**, 817–22.
20. Zhang, Z. and Yuen, G. Y. 2000, Effect of culture fluids and preincubation of chitinase production on biocontrol of *Bipolaris* leaf spot by *Stenotrophomonas maltophila* C3. *Biol. Control* **18**, 277–86.
21. Christensen, P. and Cook, F. D. 1978, *Lysobacter*, a new genus of nonfruiting, gliding bacteria with high base ratio. *Int. J. Syst. Bacteriol.*, **28**, 367–93.
22. Tindall, B. J. and Euzéby, J. P. 2006, *Lysobacter enzymogenes* subsp. *enzymogenes* Christensen and Cook 1978, *L. enzymogenes* subsp. *cookii* Christensen 1978 and *Streptococcus casseliflavus* (Mundt and Graham 1968) Vaughan *et al.* 1978 should have been cited in the approved lists of bacterial names. Request for an opinion. *Int. J. Syst. Evolv. Microbiol.*, **56**, 2707–9.
23. Epstein, D. M. and Wensink, P. C. 1988, The α-lytic protease gene of *Lysobacter enzymogenes*. The nucleotide sequence predicts a large prepro-peptide with homology to pro-peptides of other chymotrypsin-like enzymes. *J. Biol. Chem.* **263**, 16586–90.
24. Chohnan, S., Nonaka, J., Teramoto, K. et al. 2002, *Lysobacter* strain with high lysyl endopeptidase production. *FEMS Microbiol. Lett.* **213**, 13–20.
25. Chohnan, S., Shiraki, K., Yokoya, K. et al. 2004, A second lysine-specific serine protease from *Lysobacter* sp. strain IB-9374. *J. Bacteriol.* **186**, 5093–100.
26. Yuen, G. Y., Steadman, J. R., Lindgren, D.T., Schaff, D. and Jochum, C. 2001, Bean rust biological control using bacterial agents. *Crop Protect.* **20**, 395–402.
27. Folman, L. B., Postma, J., and van Veen, J. A., 2003, Characterization of *Lysobacter enzymogenes* (Christensen and Cook 1978) strain 3.1T8, a powerful antagonist of fungal diseases of cucumber. *Microbiol. Res.*, **158**, 107–15.
28. Chen, J., Moore, W. H., Yuen, G. Y., Kobayashi, D. and Caswell-Chen, E. P. 2006, Influence of *Lysobacter enzymogenes* strain C3 on Nematodes. *J. Nematol.*, **38**, 233–9.
29. Mercer, C. F., Greenwood, D. R. and Grant, J. L. 1992. Effect of plant and microbial chitinases on the eggs and juveniles of *Meloidogyne hapla* Chitiwood. *Nematologica* **38**, 227–36.
30. Zhang, Z. and Yuen, G. Y. 2000, The role of chitinase production by *Stenotrophomonas maltophilia* strain C3 in biological control of *Bipolaris sorokiniana*. *Phytopathology* **90**, 384–9.
31. Zhang, Z., Yuen, G. Y., Sarath, G. and Penheiter, A.R. 2001, Chitinases from the plant disease biocontrol agent, *Stenotrophomonas maltophilia* C3. *Phytopathology* **91**, 204–10.
32. Kobayashi, D.Y. and El-Barrad, N.E.-H. 1996, Selection of bacterial antagonists using enrichment cultures for the control of summer patch disease in Kentucky Bluegrass. *Cur. Microbiol.* **32**, 106–11.
33. Palmubo, J.D., Sullivan, R.F. and Kobayashi, D.Y. 2003, Molecular characterization and expression in *Escherichia coli* of three β-1,3-glucanase genes from *Lysobacter enzymogenes* strain N4-7. *J. Bacteriol.*, **185**, 4362–70.
34. Hashizume, H., Igarashi, M., Hattori, S., Hori, M., Hamada, M. and Takeuchi, T. 2001, Tripropeptins, novel antimicrobial agents produced by *Lysobacter* sp., I. Taxonomy, isolation and biological activities. *J. Antibiotics* **54**, 1054–9.
35. Hashizume, H., Igarashi, M., Sawa, R., Adachi, H., Nishimura, Y. and Akamatsu, Y. 2008, A new type of tripropeptin with anteiso-branched chain fatty acid from *Lysobacter* sp. BMK333-48F3. *J. Antibiotics* **61**, 577–82.
36. Hamamoto, H., Urai, M., Ishii, K., et al. 2015. Lysocin E is a new antibiotic that targets menaquinone in the bacterial membrane. *Nat. Chem. Biol.* **11**, 127–33.
37. de Brujin, I., Cheng, X., de Jager, V., et al. 2015. Comparative genomics and metabolic profiling of the genus *Lysobacter*. *BMC Genomics*, **16**, 991.
38. Gordon, D., Desmarais, C. and Green, P. 2001. Automated finishing with autofinish. *Genome Res.* **11**, 614–25.
39. Jong, A. D., Pietersma, H., Cordes, M., Kuipers, O., Kok, J. 2012, PePPER: a webserver for prediction of prokaryote promoter elements and regulons. *BMC Genomics*, **13**, 299.
40. Rawlings, N. D., Waller, M., Barrett, A. J. and Bateman, A. 2014, *MEROPS*: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.*, **42**, D503–509.
41. Uchiyama, I. 2006, Hierarchical clustering algorithm for comprehensive orthologous-domain classification in multiple genomes. *Nucleic Acids Res.*, **34**, 647–58.
42. Uchiyama, I., Mihara, M., Nishide, H. and Chiba, H. 2015, MBGD update 2015: microbial genome database for flexible ortholog analysis utilizing a diverse set of genomic data. *Nucleic Acids Res.*, **43**, D270–6.
43. Uchiyama, I. 2008, Multiple genome alignment for identifying the core structure among moderately related microbial genomes. *BMC Genomics* **9**, 515.
44. Ansari, M.Z., Yadav, G., Gokhale, R.S. and Mohanty, D. 2004, NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res.*, **32**, W405–13.
45. Tae, H., Kong, E.B. and Park, K. 2007, ASMPKS: an analysis system for modular polyketide synthases. *BMC Bioinformatics*. **8**, 327.
46. Bachmann, B.O. and Ravel J.. 2009, Methods for *in silico* prediction of microbial secondary metabolic pathways from DNA sequence data. *Methods Enzymol.*, **458**, 181–217.
47. Medema, M.H., Blin, K., Cimermancic, P., et al. 2011, antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.*, **39**, W339–46.
48. Takami, H., Taniguchi, T., Arai, W., Takemoto, K., Moriya, Y. and Goto, S. 2016, An automated system for evaluation of the potential functionome: MAPLE version 2.1.0. *DNA Res.*, **23**, 467–475.
49. Takami, H., Taniguchi, T., Moriya, Y., Kuwahara, T., Kanehisa, M. and Goto, S. 2012, Evaluation method for the potential functionome harbored in the genome and metagenome. *BMC Genomics*, **13**, 699.
50. Braun, V. 1997, Avoidance of iron toxicity through regulation of bacterial iron transport. *Biol. Chem.* **378**, 779–786.
51. Noinaji, N., Guillier, M., Barnard, T.J., and Buchanan, S.K. 2010, TonB-dependent transporters: regulation, structure, and function. *Annu. Rev. Microbiol.* **64**, 43–60.
52. Sonenshein, A.L., Hoch, J.A. and Losic, R (eds.), *Bacillus subtilis and Other Gram-Positive Bacteria*. ASM Press, Washington, DC.
53. Walsh, C.T. 2003, *Antibiotics; Actions, Origins, Resistance*. ASM Press, Washington, DC.
54. Xie, Y., Wright, S., Shen, Y. and Du, L. 2012, Bioactive natural products from *Lysobacter*. *Nat. Prod. Pep.* **29**, 1277–87.
55. Hou, J., Robbel, L. and Marahiel, M.A. 2011, Identification and characterization of the lysobactin biosynthetic gene cluster reveals mechanistic insights into an unusual termination module architecture. *Chem. Biol.* **18**, 655–64.
56. Kobayashi, D.Y., Reedy, R.M., Palumbo, J.D., Zhou, J.-M. and Yuen, G.Y. 2005, A *clp* gene homologue belonging to the *crp* gene family globally regulates lytic enzyme production, antimicrobial activity, and biological control activity expressed by *Lysobacter enzymogenes* strain C3. *Appl. Environ. Microbiol.* **71**, 261–9.
57. Wang, Y., Zhao, Y., Zhan, J., et al. 2014, Transcriptomic analysis reveals new regulatory roles of Clp signaling in secondary metabolite synthesis and surface motility in *Lysobacter enzymogenes* OH11. *Appl. Microbiol. Biotechnol.*, **98**, 9009–20.
58. Han, Y., Wang, Y., Tombosa, S., et al. 2015, Identification of a small molecule signaling factor that regulates the biosynthesis of the antifungal polycyclic tetramate macrolactam HSAF in *Lysobacter enzymogenes*. *Appl. Microbiol. Biotechnol.* **99**, 801–11.