SCIENCE CHINA
Life Sciences

• RESEARCH PAPERS •

September 2011 Vol.54 No.9: 835–840
doi: 10.1007/s11427-011-4211-9

A complete genome assembly of *Glaciecola mesophila* sp. nov. sequenced by using BIGIS-4 sequencer system

YUAN LiNa[1†], REN LuFeng[1†], LI YunTao[2], HAN WeiJing[2], YU Yong[3], CHU YaNan[1], LIU GuiMing[1], YU Dan[1], TENG MingJing[1], WANG Liang[1], WANG XuMin[1], ZHOU XiaoGuang[2], YU YuDe[2*] & YU Jun[1*]

[1]*Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, China;*
[2]*Institute of Semiconductor, Chinese Academy of Sciences, Beijing 100083, China;*
[3]*SOA Key Laboratory for Polar Science, Polar Research Institute of China, Shanghai 200136, China*

Received May 9, 2011; accepted June 10, 2011; published online August 2, 2011

Using a pyrosequencing-based custom-made sequencer BIGIS-4, we sequenced a Gram-negative bacterium *Glaciecola mesophila* sp. nov. (*Gmn*) isolated from marine invertebrate specimens. We generated 152043 sequencing reads with a mean high-quality length of 406 bp, and assembled them using the BIGIS-4 post-processing module. No systematic low-quality data was detected beyond expected homopolymer-derived errors. The assembled *Gmn* genome is 5144318 bp in length and harbors 4303 annotated genes. A large number of metabolic genes correspond to various nutrients from surface marine invertebrates. Its abundant cold-tolerant and cellular signaling and related genes reveal a fundamental adaptation to low-temperature marine environment.

***Glaciecola mesophila* sp. nov., BIGIS-4 sequencer system, marine environment adaptation**

Citation: Yuan L N, Ren L F, Li Y T, *et al*. A complete genome assembly of *Glaciecola mesophila* sp. nov. sequenced by using BIGIS-4 sequencer system. Sci China Life Sci, 2011, 54: 835–840, doi: 10.1007/s11427-011-4211-9

The genus *Glaciecola* accommodates Gram-negative, motile, aerobic, psychrophilic, pigmented, and slightly halophilic bacteria, widely spread in marine water, ice, sediment, and marine invertebrates [1–8]. Most *Glaciecola* strains produce oxidase, catalases, esterases, amylases, agaropectinases, α/β galactosidases, xylanases, and other enzymes degrading large molecules and bacterial polysaccharides. These biological properties indicate that *Glaciecola* may have important roles in the material cycling of marine environments. Therefore, *Glaciecola* species are excellent model organisms for the investigation of ecological evolution in marine ecosystems, even in crude oil-contaminated sea water [9]. *Glaciecola mesophila* sp. nov. (*Gmn*) is a novel model strain, isolated from internal liquor of a specimen of the ascidian *Halocynthia aurantium* from Troitsa Bay, the Sea of Japan, Russia [1]. It grows optimally at 25°C in the presence of 3% (w/v) NaCl and at pH 7.5 but does not survive below 10°C and above 35°C. It grows poorly in 1% NaCl and shows inhibited growth in 6% NaCl. It does not hydrolyze urea, gelatin, or casein; weakly produces acids from carbohydrates, and does not utilize L-arabinose, citrate, phenylacetate, adipate, L-phenylalanine, L-histidine, L-ornithine, L-threonine and putrescine. Guo *et al*. [10] cloned and expressed a cold-active and salt-tolerant enzyme, xylanase, from *G. mesophila* KMM 241. The xylanase efficiently hydrolyzes xylo-oligosaccharides and xylan into xylobiose and xylotriose without producing xylose. The xylanase retained more activity and catalytic efficiency at 4°C and remained active under high sodium ion concentration. The genomic sequence of *Glaciecola* would provide a step towards the

†Contributed equally to this work
*Corresponding author (email: yudeyu@semi.ac.cn; junyu@big.ac.cn)

better understanding of marine microorganism's adaptation to their unique environments at the genomic level.

BIGIS-4 is a scalable, highly integrative system for determining nucleotide sequences based on the pyrosequencing chemistry. It comprises an optics module, a fluidics module, a computer module, four reaction modules, and a post-processing module. The reaction modules are in four separate identical chambers and share one fluidics system. Capable of handling variable amounts of samples, one to four reaction module(s) can to be launched to accommodate the throughput requirement. The apparatus uses a fiber-optic slide of individual wells and is able to sequence 560 million bases in one run. BIGIS-4 generates amplified sequencing data from a four-step reaction, supported by four enzymes with dNTPs flowing in one by one sequentially, based on the pyrosequencing rationale [11]. The chemically generated light signal is monitored by CCD imaging *in situ* and transferred to the post-processing module. Thus, BIGIS-4 is a next generation sequencing system with a throughput paralleling 454 and with a read length much superior to the Life Technologies' SOLiD system and Illumina's GAII system. Here, we present a genome assembly of *Gmn* based on data acquired with the BIGIS-4 one-reaction module.

1 Materials and methods

1.1 Cultivation and genome DNA preparation

The strains were cultured in a medium prepared from an artificial seawater base ($MgCl_2$ 5 g L^{-1}, $MgSO_4$ 2 g L^{-1}, $CaCl_2$ 0.5 g L^{-1}, KCl 1 g L^{-1}, peptone 5 g L^{-1}, yeast extract 0.1 g L^{-1}) supplemented with 3% (w/v) NaCl and adjusted to pH 7.5 with 1 N KOH, at 25°C for 2 d.

1.2 Genome sequencing

DNA purification, library preparation and amplification were carried out according to General Library Preparation Method and emPCR Amplification Method of 454. The bead-immobilized, amplified DNA library was loaded into micro-wells in the BIGIS-4 Seq-plate, along with enzyme (sulfurylase and luciferase) beads and DNA polymerase by centrifugation. At the same time, the pre-wash buffer filled reagent cassette was ready for a one-hour prewash. After pre-washing, the pre-wash buffer was replaced with BIGIS-4 sequencing reagent kit (including dATP, dTTP, dCTP, dGTP, substrates, ATP, apyrase, and buffer). The used Seq-plate was removed, and the beads-loaded plate was fixed in the reaction chamber. The plate back was set against the fiber face of the CCD camera, which forms a fully dark region for chemiluminescent light signal capture. Raw data of a sequencing run were transferred to the BIGIS-4 Post-Run software module for base calling and data processing.

1.3 Sequence assembly

The sequencing reads were first assembled with BIGIS-4 assembler. Gaps between assembled contigs were closed by PCR walking (for gaps<2 kb) and by shotgun sequencing (for gaps>2 kb). A hybrid assembly of BIGIS-4 and Sanger reads was performed using Consed [12], coupled with intensive manual alignment and visual inspection. The prediction of coding sequences was performed with Glimmer3 [13]. All predicted proteins were searched against a non-redundant protein database (nr, NCBI) using BLASTP with an E-value cutoff of 1×10^{-10}. Protein signatures were recognized by InterProScan [14]. All predicted proteins were searched for against the COG database. In addition to coding genes, tRNAs were annotated using tRNAscan-SE [15], and rRNAs were detected by the RNAmmer 1.2 server (http://www.cbs.dtu.dk/services/RNAmmer/). Insertion sequences were detected by ISFinder [16]. Comparative genomics analysis was carried out using MUMmer [17]. For the refinement of input sequences and annotation, several custom-designed, perl-based scripts were also developed.

2 Results

2.1 Genome assembly based on BIGIS-4 reads

The BIGIS-4 one-reaction module yielded 152043 high quality reads. Figure 1 illustrates the distribution of read length. The mean and median lengths were 406 and 442 bp, respectively. The estimated accuracy was 99.5%, and the homopolymer-derived error rate was as expected from the chemistry principle used. No systematic low-quality data were detected. The reads were assembled into 157 contigs, with an N50 size of 61174 bp. Nine gaps were greater than 2 kb long, and the mean Sanger sequencing reads for closure were 485 bp. Our *in silico* evaluation showed that the final genome was evenly covered by BIGIS-4 reads (Figure 2) at a mean sequencing depth of 11.9×. The BIGIS-4 reads recovered 99.4% of the final genome. The high-coverage regions (>60×) contain mainly transposases, multicopy hypothetical proteins in the genome, other than rRNA operons, and include 11 mobile element-related genes and six from insertion sequence (IS) IS4 family. On the other hand, genes that were absent from the coverage graph with assigned functions were ABC transporter-related, DNA-binding transcriptional activators, response regulators, and peptidase-like proteins.

2.2 General features of *Gmn* genome

The *Gmn* genome consists of a circular chromosome of 5144318 bp in length (Figure 3). We identified 4528 predicted protein-coding, 59 tRNA, and 15 rRNA genes. The 4303 annotated genes comprise 3220 orthologs assigned to known functions, 1043 hypothetical proteins, and 40 mobile

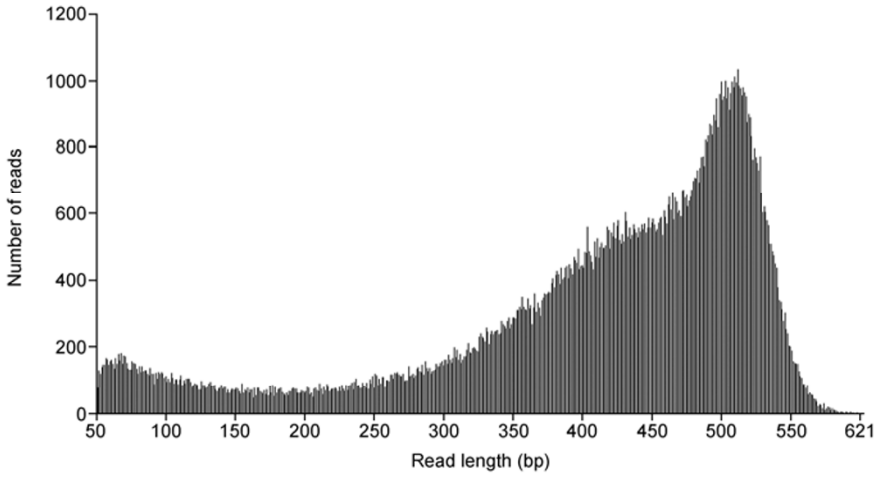

Figure 1 Read length distribution from the BIGIS-4 one-reaction module.

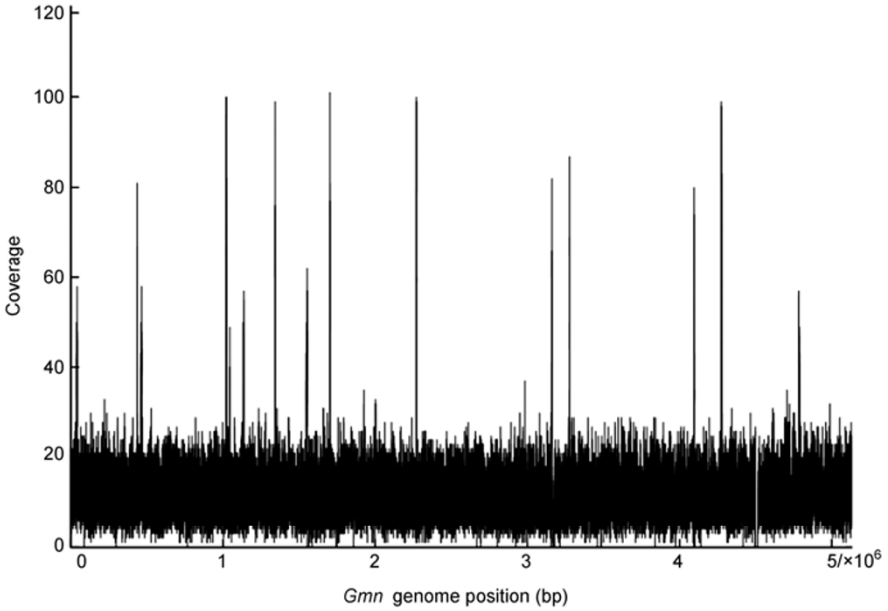

Figure 2 BIGIS-4 reads coverage on the *Gmn* genome sequence using Mosaik [18].

elements, such as integrases (8), transposases (26), or phage-related (6). The coding density is 86% and intergenic spacers have a mean length of 166 bp. The GC content of the assembled sequence is 44.6%.

We identified 10 complete copies of IS elements in the assembly. These IS elements are classified into two families (IS4 and IS110); the IS4 family (seven complete copies/isoforms) was predominant.

The annotated genes include a variety of metabolic genes for sugar, glucose, sorbosone, alpha-L-fucoside, galactoside, fucose, mannose, and N-acetylgalactosamine. There is a beta-1,4-xylanase associated with cold adaption of bacteria as well as a gene involved in defense against osmotic stress. Genes for ABC-type sugar transporters and TRAP-type C4-dicarboxylate transporters, respectively, are aligned adjacently and may both function as operons. The genome assembly also contains abundant two-component signal transduction proteins, transcriptional regulators, heat shock response proteins, flagellar/pilus assembly proteins, chaperone proteins, and secretory proteins, enabling the strain to efficiently respond to diverse environmental changes.

2.3 Comparative genomics

For comparative genomic analysis, we compared our sequence assembly to those of *Pseudoalteromonas atlantica* T6c and *Alteromonas macleodii* 'Deep ecotype' from the NCBI database. The results showed that *Gmn* is phylogenetically close to *P. atlantica* T6c, based on sequence similarity (Figure 4), which is consistent with a 16S rRNA-

based analysis (Figure 5).

The synteny between *Gmn* and *P. atlantica* T6c is strong, and yet there are only a few extended homologous regions scattered over different locations. Of 4303 annotated genes of *Gmn* genome, 3968 have best hits in the *P. atlantica* T6c sequence, but only 13 hits in the *A. macleodii* 'Deep ecotype' sequence were identified. Among the genes missing from *Gmn* genome when compared with *P. atlantica* T6c, 17 are integrases or transposases, which are candidates for horizontal gene transfer, 62 are hypothetical and eight have multiple copies in the genome. In contrast, *A. macleodii* 'Deep ecotype' and *Gmn* share many conserved genes with various functions, and genes missing from the former are mainly for metabolism; most of the genes

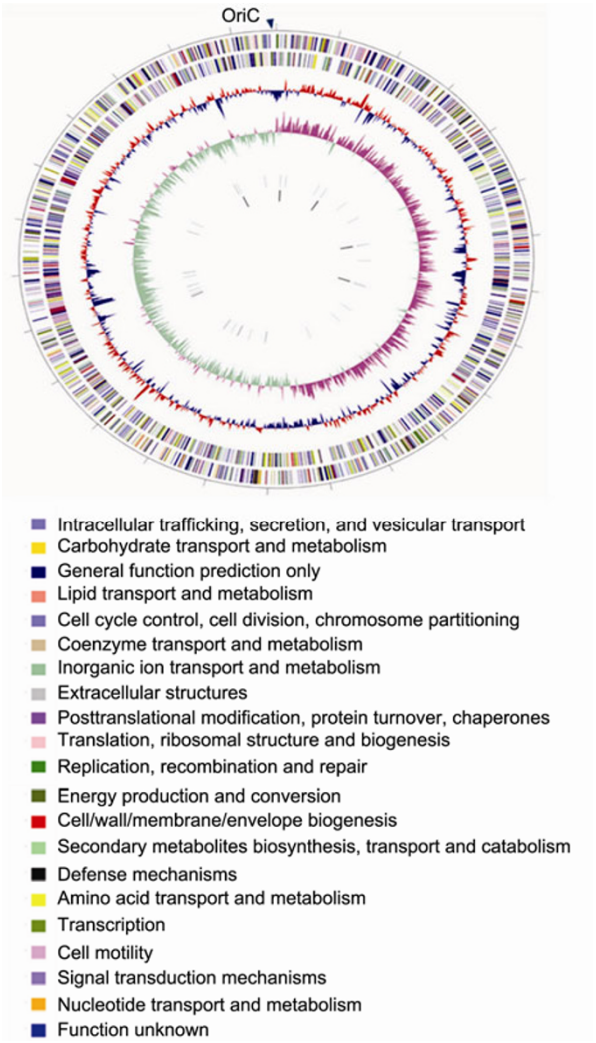

Figure 3 Circular representation of the *Gmn* genome. From the outside in: Circle 1, scale bar in kb. Circles 2 and 3, strand-dependent depiction of all CDS. Circle 4, GC content along the genome (10 kb window in 1 kb step), red (>44.6%), blue (<44.6%). Circle 5, GC skew (G–C/G+C, 10 kb window in 1 kb step), red-purple (GC skew>0), green (GC skew<0). Circle 6, tRNAs. Circle 7, rRNAs. The OriC site was predicted with GC skew, where the replication terminus is located almost diametrically opposite the origin of replication and is accompanied by a sharp transition in the GC skew.

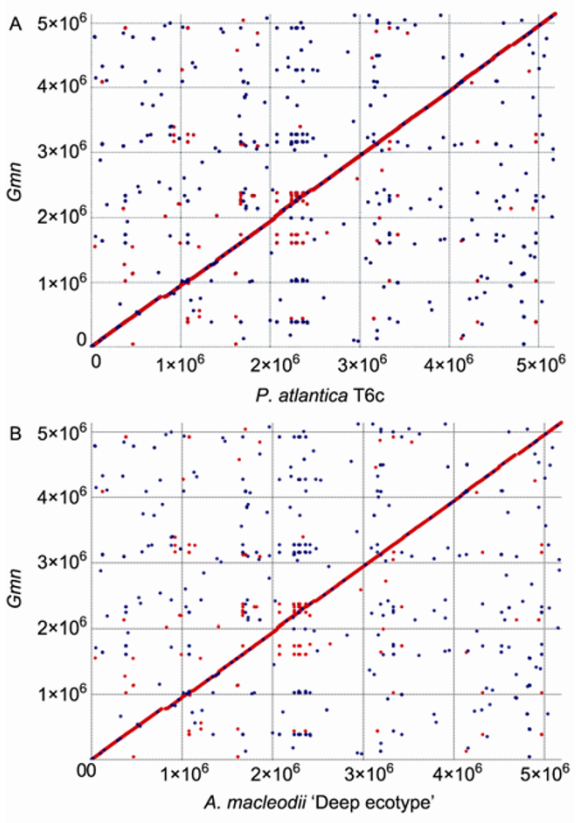

Figure 4 Whole genome alignment of *Gmn* vs. *P. atlantica* T6c (A) and *Gmn* vs. *Alteromonas macleodii* 'Deep ecotype' (B) using MUMmer. The reference sequence is laid across the *x*-axis, while the query sequence is on the *y*-axis. The colored line or dot is plotted wherever the two sequences agree. The forward matches are displayed in red, while the reverse matches are displayed in blue. A single red line would go from the bottom left to the top right when the two sequences were perfectly identical.

(69.8%) unique to *A. macleodii* 'Deep ecotype' encode hypothetical proteins, and the rest do not show any biased distribution. *A. macleodii* 'Deep ecotype' also contains more transposable elements and large numbers of IS elements [19].

2.4 Environmental adaption

We identified a gene encoding a cold-active and salt-tolerance enzyme family, endo-beta-xylanase, in our assembly. This family contains 10 endo-beta-xylanases which degrade linear polysaccharide beta-1,4-xylan into xylose [20]. The xylanases were found in *P. atlantica* T6c, but not in *A. macleodii* 'Deep ecotype'. This taxonomic affiliation between *Gmn* and *P. atlantica T6c* was also shared by the GroE chaperons, which showed identical amino acid sequences. The GroE chaperons comprise two co-operating proteins, GroES and GroEL, belonging to a class of chaperone proteins highly conserved during evolution. The two genes are arranged in *Gmn* genome in the order *groES- groEL*, separated by a 35 bp intergenic region. GroEL showed remarkably high similarity (with query coverage >95% and

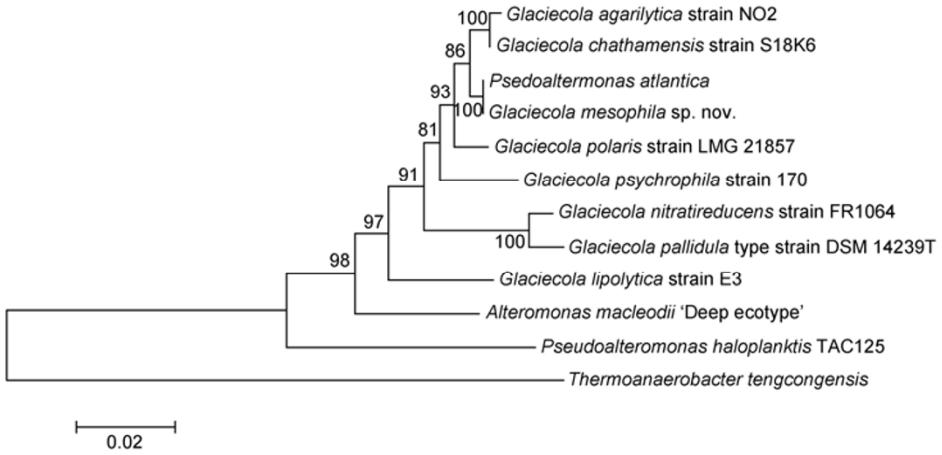

Figure 5 Phylogenetic tree constructed by the neighbor-joining method using the MEGA4 software. The node labels are bootstrap values. The 16S rDNA sequences of *Glaciecola* and related taxa were retrieved from GenBank databases and were aligned with ClustalW program. *Thermoanaerobacter tengcongensis* was used as an outgroup.

identities >80%) with those from psychrophilic, mesophilic, thermophilic, and other bacteria (data not shown).

3 Discussion

We report a complete genome sequence assembly of *Gmn*, a novel marine strain isolated from a marine invertebrate, using our custom-made BIGIS-4 system. On the basis of phenotypic, chemotaxonomic, and phylogenetic data, we conclude that *Gmn* represents a novel species of the genus *Glaciecola*. For genomic comparison, genomes of two related strains were obtained. *A. macleodii* 'Deep ecotype' is a common deep sea isolate obtained from depths of 1000 m in the South Adriatic Sea basin [21], and has been reported to be phylogenetically close to *Glaciecola* [1]. *P. atlantica* T6c is a marine bacterium isolated from a biofilm from San Diego Bay in California. The genus *Pseudoalteromonas* was first proposed by Gauthier *et al*. [22] as a new genus split from *Alteromonas*. Whole genome alignments showed that *P. atlantica* T6c is the closest relative of *Gmn*. In fact, on the basis of 16S rRNA sequences, *Gmn* and *P. atlantica* T6c are also mutually the closest neighbors. It might represent a significant fraction of the microbiota in marine invertebrates. When protein similarity was scored in a genome-wide fashion, more than 90% of *Gmn* genes have extensive similarity to those of *P. atlantica* T6c. *A. macleodii* 'Deep ecotype' represents a deep sea ecotype. As expected, the number of genes shared, and average similarity, are much lower for *A. macleodii* 'Deep ecotype'. In contrast to *A. macleodii* 'Deep ecotype', *Gmn* genome contains abundant genes acting as metabolic genes, degrading more sugars and amino acids, which might indicate a wider range of nutrition available compared with that in the deep sea.

A gene encoding an endo-beta-1,4-xylanase, which is often associated with high tolerance to cold environments and high salt concentrations, was identified in the *Gmn* genome. Collins *et al*. [20] had reported that the optimal temperatures for the activity of psychrophilic xylanases are apparently lower than that of the mesophilic reference, which provides evidence of cold environment adaptation. Based on physiological experiments, *A. macleodii* type strain can survive at 37°C even 40°C, and also has a higher salt-tolerance at 10% NaCl, a concentration at which neither *Gmn* nor *P. atlantica* T6c is able to survive [1]. It is suggested there might be a discrepant mechanism for temperature and salt stress, which supports these marine strains thrive in their habitats. Another genetic element that is likely to be a major source for adaptation to marine environments is the chaperons GroE. GroE is involved in productive folding of proteins which helps bacteria to cope with a variety of stresses, including heat shock, NaCl, and SDS [23]. GroE is predicted to be highly expressed when exposed to those stresses [24], especially from deep-sea genomes [25]. Taken together with abundant two-component signal transduction proteins, transcriptional regulators in the genome, and the related cellular processes and signaling, the proteins enable the strain to effectively respond to extra- and intra-cellular stimuli.

The genus *Glaciecola* accommodates strains that thrive in a variety of marine niches, with variable temperatures and nutrition demands. Correspondingly, various genomic adaption mechanisms have formed. The identification of strains with similarly basic genomics would result in more convincing evidence for adaptation. Further genomic sequences are expected to appear in the near future, facilitating a systemic investigation of genomic evolution for adaptation to marine environments.

This work was supported by the Chinese Academy of Sciences Scientific Research Equipment (Grant No. YZ200823), the Institutional Director's Initiative Fund awarded to Yu Jun, the National Natural Science Founda-

tion of China (Grant Nos. 61007033, 30971610, and 40906097), and the Institutional Initiative Fund awarded to Wang XuMin.

1 Romanenko L A, Zhukova N V, Rohde M, *et al*. *Glaciecola mesophila* sp. nov., a novel marine agar-digesting bacterium. Int J Syst Evol Microbiol, 2003, 53: 647–651
2 Zhang D C, Yu Y, Chen B, *et al*. *Glaciecola psychrophila* sp. nov., a novel psychrophilic bacterium isolated from the Arctic. Int J Syst Evol Microbiol, 2006, 56: 2867–2869
3 Van Trappen S, Tan T L, Yang J, *et al*. *Glaciecola polaris* sp. nov., a novel budding and prosthecate bacterium from the Arctic Ocean, and emended description of the genus *Glaciecola.* Int J Syst Evol Microbiol, 2004, 54: 1765–1771
4 Baik K S, Park Y D, Seong C N, *et al*. *Glaciecola nitratireducens* sp. nov., isolated from seawater. Int J Syst Evol Microbiol, 2006, 56: 2185–2188
5 Chen L P, Xu H Y, Fu S Z, *et al*. *Glaciecola lipolytica* sp. nov., isolated from seawater near Tianjin city, China. Int J Syst Evol Microbiol, 2009, 59: 73–76
6 Matsuyama H, Hirabayashi T, Kasahara H, *et al*. *Glaciecola chathamensis* sp. nov., a novel marine polysaccharide-producing bacterium. Int J Syst Evol Microbiol, 2006, 56: 2883–2886
7 Zhang Y J, Zhang X Y, Mi Z H, *et al*. *Glaciecola arctica* sp. nov., isolated from Arctic marine sediment. Int J Syst Evol Microbiol, 2010
8 Yong J J, Park S J, Kim H J, *et al*. *Glaciecola agarilytica* sp. nov., an agar-digesting marine bacterium from the East Sea, Korea. Int J Syst Evol Microbiol, 2007, 57: 951–953
9 Prabagaran S R, Manorama R, Delille D, *et al*. Predominance of *Roseobacter*, *Sulfitobacter*, *Glaciecola* and *Psychrobacter* in seawater collected off Ushuaia, Argentina, Sub-Antarctica. FEMS Microbiol Ecol, 2007, 59: 342–355
10 Guo B, Chen X L, Sun C Y, *et al*. Gene cloning, expression and characterization of a new cold-active and salt-tolerant endo-beta-1,4-xylanase from marine *Glaciecola mesophila* KMM 241. Appl Microbiol Biotechnol, 2009, 84: 1107–1115
11 Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. Science, 1998, 281: 363–365
12 Gordon D. Viewing and editing assembled sequences using Consed. Curr Protoc Bioinformatics, 2003, Chapter 11: Unit11.2
13 Delcher A L, Bratke K A, Powers E C, *et al*. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics, 2007, 23: 673–679
14 Zdobnov E M, Apweiler R. InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics (Oxford, England), 2001, 17: 847–848
15 Schattner P, Brooks A N, Lowe T M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res, 2005, 33: W686–W689
16 Siguier P, Perochon J, Lestrade L, *et al*. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res, 2006, 34: D32–D36
17 Kurtz S, Phillippy A, Delcher A L, *et al*. Versatile and open software for comparing large genomes. Genome Biol, 2004, 5: R12
18 Bott O J, Bergmann J, Hoffmann I, *et al*. Analysis and specification of telemedical systems using modelling and simulation: the MOSAIK-M approach. Stud Health Technol Inform, 2005, 116: 503–508
19 Ivars-Martinez E, Martin-Cuadrado A B, D'Auria G, *et al*. Comparative genomics of two ecotypes of the marine planktonic copiotroph *Alteromonas macleodii* suggests alternative lifestyles associated with different kinds of particulate organic matter. ISME J, 2008, 2: 1194–1212
20 Collins T, Gerday C, Feller G. Xylanases, xylanase families and extremophilic xylanases. FEMS Microbiol Rev, 2005, 29: 3–23
21 Tosco A, Birolo L, Madonna S, *et al*. GroEL from the psychrophilic bacterium *Pseudoalteromonas haloplanktis* TAC 125: molecular characterization and gene cloning. Extremophiles, 2003, 7: 17–28
22 Yamauchi S, Okuyama H, Morita E H, *et al*. Gene structure and transcriptional regulation specific to the groESL operon from the psychrophilic bacterium *Colwellia maris.* Arch Microbiol, 2003, 180: 272–278
23 Xu K, Ma B G. Comparative analysis of predicted gene expression among deep-sea genomes. Gene, 2007, 397: 136–142
24 Lopez-Lopez A, Bartual S G, Stal L, *et al*. Genetic analysis of housekeeping genes reveals a deep-sea ecotype of *Alteromonas macleodii* in the Mediterranean Sea. Environ Microbiol, 2005, 7: 649–659
25 Gauthier G, Gauthier M, Christen R. Phylogenetic analysis of the genera *Alteromonas*, *Shewanella*, and *Moritella* using genes coding for small-subunit rRNA sequences and division of the genus *Alteromonas* into two genera, *Alteromonas* (emended) and *Pseudoalteromonas* gen. nov., and proposal of twelve new species combinations. Int J Syst Bacteriol, 1995, 45: 755–761