Complete genome sequence of a marine bacterium with two
chromosomes, *Pseudoalteromonas translucida* KMM 520^T

Jin-Cheng Rong^a, Min Liu^a, Yi Li^a, Tian-Yong Sun^a, Xiu-Hua Pang^a, Qi-Long Qin^{a,b},
Xiu-Lan Chen^{a,b}, Bin-Bin Xie^{a,b,*}

^a State Key Laboratory of Microbial Technology, Marine Biotechnology Research Center,
Shandong University, Jinan 250100, China

^b Institute of Marine Science and Technology, Shandong University, Jinan 250100, China

* Corresponding author. E-mail address: xbb@sdu.edu.cn (B. Xie).

Abstract

Bacteria with multiple chromosomes provide new insights into the evolution of multipartite genome structures and bacterial survival strategies. In this study, we report the complete genome sequence of *Pseudoalteromonas translucida* KMM 520^T, which contains two circular chromosomes and comprises 4,147,593 bp with a mean G + C content of 40.1%. The two chromosomes have similar G + C contents and similar percentages of coding regions. Chromosome II of *P. translucida* possesses a plasmid-type replication initiator protein (RepA), which indicated that chromosome II is probably originated from a plasmid. COG functional categories revealed that the two chromosomes have divergent distributions of functional categories, which indicated that they bear different responsibilities for cellular functions. The complete genome sequence of *P. translucida* contributes to a better understanding of the origin and

evolution of the additional chromosome and the physiology of *Pseudoalteromonas* genus.

Keywords: Multiple chromosomes; *Pseudoalteromonas*; COG functional categories

1 Introduction

For a fairly long period of time, bacteria were thought to contain only one chromosome. The era of genomics shed light on the fact that some bacteria possess one or more additional chromosomes and these bacteria are mainly limited to the genus of *Brucella*, *Agrobacterium*, *Burkholderia*, *Ralstonia*, *Vibrio* and *Leptospira* (Egan et al., 2005; Jha et al., 2012; Prozorov, 2008). Bacteria with multiple chromosomes provide opportunities to understand the origin and evolution of multipartite genome structures, the coordination of replication and segregation of the individual chromosomes and their ecological functions (Bavishi et al., 2010; Egan et al., 2005; Ochman, 2002). Currently, complete genome sequences of only a small number of multichromosomal bacteria are available. The origin, evolution, mechanisms of replication and ecological functions of multichromosomal bacteria remain largely unknown.

Pseudoalteromonas, belonging to *Alteromonadales* of *Gammaproteobacteria*, is a genus of Gram-negative, heterotrophic, flagellated and **bioactive compound producing bacteria** widespread in the marine environment (Bowman, 2007; Gauthier et al., 1995). The genus currently contains 39 species. Very limited complete genomes of the genus showed that some *Pseudoalteromonas* genomes contain two chromosomes (Medigue

et al., 2005; Qin et al., 2011). Here we sequenced the complete genome of *Pseudoalteromonas translucida* KMM 520^T (KMM 520^T) and discovered that KMM 520^T also contains two circular chromosomes. Strain KMM 520^T was isolated from sea water of Troitza Bay in Sea of Japan at a water depth of 5-8 m (Ivanova et al., 2002). The strain are deposited in Collection of Marine Micro-organisms (Vladivostok, Russian Federation) with the accession number KMM 520^T (= DSM14402^T = LMG 19696^T = ATCC BAA-315^T). In this study, we report the first genome sequence of *P. translucida* to provide insights into the evolution of the additional chromosome and the physiology of *Pseudoalteromonas* genus.

2 Data description

General features of strain KMM 520^T were summarized in Table 1. Strain KMM 520^T was cultured in marine broth 2216 (MB, BD) at 25°C. The total genomic DNA was extracted using PowerMaxTM Soil DNA Isolation Kit (MO BIO, USA). The genome sequence of KMM 520^T was determined using a combination of Solexa sequencing (HiSeq2000 platform, Illumina, USA) and Sanger sequencing. Solexa paired-end libraries of different insert lengths, 500 bp, 2000 bp and 5000 bp, were constructed. A total of 1000 Mb clean data with every read sequence of 90 bp from the Solexa sequencing were obtained, representing a 241-fold coverage. *De novo* assembly was performed using SOAPdenovo version 1.05 (Li et al., 2010). The draft genome was completed by primer walking and Sanger sequencing of PCR segments. Six gaps with a total size of 38 kb were closed by 16 PCR.

The tRNA genes were identified by tRNAscan-SE v.1.3 (Lowe and Eddy, 1997). The rRNA genes were predicted by RNAmmer v.1.2 (Lagesen et al., 2007). Gene

70 calling was performed by Glimmer v.3.0 (Delcher et al., 2007). The predicted ORFs
71 were annotated by Kyoto Encyclopedia of Genes and Genomes (KEGG) with the
72 BBH method (Moriya et al., 2007). Functional categories of ORFs were performed by
73 BLASTP search against Clusters of Orthologous Groups (COG) database (Tatusov et
74 al., 2001) with E-value cutoff of 1E-5 and identity cutoff of 30%.

75

76 Table 1. General features of *Pseudoalteromonas translucida* KMM 520^T and MIGS
77 mandatory information.

Items	Description
General features	
Gram stain	Negative
Cell shape	Rod
Pigmentation	Non-pigmented
Temperature	4–30°C
Salinity	1 to 6%
pH range	6.0 to 10.0
Motility	Motile with bipolar flagella
MIGS data	
Investigation_type	Bacteria_archaea
Project_name	Genome sequence of <i>Pseudoalteromonas translucida</i>
Lat_lon	NA
Depth	5-8 m
Geo_loc_name	Russia:Troitza Bay
Collection_date	1997-07
Env_biome	bay [ENVO:00000032]
Env_feature	coastal sea water [ENVO:00002150]
Env_material	water [ENVO:00002006]
Env_package	sea water [ENVO:00002149]
Num_replicons	2
Extrachrom_elements	0
Estimated_size	4,147,593
Ref_biomaterial	PMID: 12361284
Source_mat_id	KMM 520 ^T ; DSM 14402 ^T ; LMG 19696 ^T ; ATCC BAA-315 ^T
Trophic_level	Chemoorganotroph
Rel_to_oxygen	Aerobic
Isol_growth_condt	PMID: 12361284
Lib_reads_seqd	11,111,112 reads
Seq_meth	ABI 3730; Illumina Hiseq 2000
Annot_source	KEGG

Finishing_strategy	Complete; 241 ×; 2 contigs
Genome assembly data	
Assembly method	SOAPdenovo v.1.05
Assembly name	PTRA_1.0
Genome coverage	241 ×
Sequencing Technology	Illumina Hiseq

78

79 The genome of KMM 520^T has a size of 4,147,593 bp with a mean G + C content
80 of 40.1% and is composed of two chromosomes (Table 2). The two chromosomes
81 have similar G + C contents (40.2% and 39.6%) and similar percentages of coding
82 regions (87.4% and 87.8%). **Meanwhile, chromosome II encodes essential genes, such**
83 **as *minCDE* operons (PTRA_b0734, PTRA_b0735 and PTRA_b0736) and *tus***
84 **(PTRA_b0790), which are not encoded on chromosome I.** Therefore, chromosome II
85 **is believed to be** a chromosome but not a plasmid. **In addition, 3140 protein-coding**
86 **sequences (CDSs) on chromosome I and 669 CDSs on chromosome II were**
87 **determined, 51.2% and 37.2% of which were annotated with defined functions**
88 **separately. Furthermore, chromosome II has a higher percentage (62.8%) of CDSs**
89 **with unknown functions** than chromosome I (48.8%). 106 tRNA genes, 9 rRNA
90 operons and 1 extra 5S rRNA gene are all annotated on chromosome I.

91

92 **Table 2. Comparison of genome features of *Pseudoalteromonas translucida* KMM**
93 **520^T, *Pseudoalteromonas haloplanktis* TAC125 and *Pseudoalteromonas* sp. SM9913.**

	<i>Pseudoalteromonas translucida</i> KMM 520 ^T		<i>Pseudoalteromonas haloplanktis</i> TAC125		<i>Pseudoalteromonas</i> sp. SM9913	
	Chr. I	Chr. II	Chr. I	Chr. II	Chr. I	Chr. II
Size (bp)	3,390,388	757,205	3,214,944	635,328	3,332,787	704,884
GC content (%)	40.2	39.6	41.0	39.3	40.4	39.8
Total CDSs	3140	669	2942	546	3077	634
CDSs with predicted functions	1608	249	1123	157	2063	422
(Percentages)	(51.2%)	(37.2%)	(38.2%)	(28.8%)	(67.0%)	(66.6%)

CDSs with unknown functions (Percentages)	1532 (48.8%)	420 (62.8)	1819 (61.8%)	389 (71.2%)	1014 (33.0%)	212 (33.4%)
Coding regions (%)	87.4	87.8	88.6	87.3	87.8	88.6
rRNA operons (16S-23S-5S)	9	0	9	0	8	0
5S rRNA (extra copies)	1	0	1	0	1	0
tRNAs	106	0	106	0	62	0

94

95 The replication origin of chromosome I (*oriI*) is in the intergenic region of gene
96 *gidA* and *mioC* while the replication origin of chromosome II (*oriII*) is located
97 between gene *repA* and *parA2* (Figure 1a). DnaA protein (PTRA_a0001), which is the
98 initiator protein of most bacterial chromosomes, initiates replication of chromosome I
99 of KMM 520^T. Meanwhile, plasmid-type RepA protein (PTRA_b0792) is responsible
100 for the replication initiation of chromosome II and indicated that chromosome II is
101 plasmid-originated.

102 It was reported that every replicon of the same genome plays a specific role in
103 cellular functions (Galardini et al., 2013). By comparing COG functional categories of
104 KMM 520^T, we discovered that chromosome I has higher percentages of genes related
105 to basic cellular functions than chromosome II (Figure 1b). These genes encoded on
106 chromosome I are mainly involved in COG categories of “J” (translation, ribosomal
107 structure and biogenesis), “H” (coenzyme transport and metabolism), “N” (cell
108 motility), “U” (trafficking, secretion, and vesicular transport), “M” (cell
109 wall/membrane/envelope biogenesis), “O” (posttranslational modification, protein
110 turnover and chaperones), “C” (energy production and conversion) and “L”
111 (replication, recombination and repair). Nevertheless, chromosome II has higher
112 percentages of genes involved in “G” (carbohydrate transport and metabolism), “E”
113 (amino acid transport and metabolism), “Q” (secondary metabolism), “K”

(transcription) and “V” (defense) (Figure 1b). For genes encoded on chromosome II and related to transcription, most genes encode transcriptional regulators with 2 genes (PTRA_b0498 and PTRA_b0650) encoding DNA-directed RNA polymerase specialized sigma subunits, 2 genes (PTRA_b0033 and PTRA_b0599) encoding Superfamily II DNA/RNA helicases and 1 gene (PTRA_b0710) encoding a Rad3-related DNA helicase. Genes involved in secondary metabolism and defense provide KMM 520^T with survival advantages, including biosynthesis of mitomycin antibiotics/polyketide fumonisin, cation/multidrug efflux pump and type I restriction-modification system. In summary, we demonstrated that chromosome I bears more responsibilities for basic cellular functions, whereas the plasmid-originated chromosome II affords KMM 520^T with some survival advantages.

Figure 1. (a) Circular representation of the *Pseudoalteromonas translucida* KMM 520^T genome. From the outside to the center: CDSs on forward strand (colored by COG categories), CDSs on reverse strand (colored by COG categories), RNA genes (tRNAs and rRNA operons), G + C content and GC skew. CDSs are depicted in different colors according to COG categories: dark grey indicates energy production and conversion (C); red indicates cell cycle control, cell division and chromosome partitioning (D); green indicates amino acid transport and metabolism (E); blue indicates nucleotide transport and metabolism (F); cyan indicates carbohydrate transport and metabolism (G); magenta indicates coenzyme transport and metabolism (H); yellow indicates lipid transport and metabolism (I); pale green indicates translation, ribosomal structure and biogenesis (J); light blue indicates transcription (K); orange indicates replication, recombination and repair (L); brown indicates Cell wall/membrane/envelope biogenesis (M); pale pink indicates cell motility (N); light grey indicates posttranslational modification, protein turnover and chaperones (O); mid red indicates inorganic ion transport and metabolism (P); light red indicates secondary metabolites biosynthesis, transport and catabolism (Q); pink indicates general function prediction only (R); dark red indicates function unknown (S); purple indicates signal transduction mechanisms (T); gold indicates intracellular trafficking, secretion, and vesicular transport (U); navy blue indicates defense mechanisms (V); black indicates hypothetical proteins. Genes surrounding the *oriI* and *oriII* sites are displayed. The solid red circles indicate the *ori* sites. (b) Comparison of COG functional categories of chromosome I and chromosome II of *P. translucida* KMM 520^T.

Nucleotide sequence accession numbers

The complete genome sequence of *P. translucida* KMM 520^T has been deposited in DDBJ/EMBL/GenBank under the accession number CP011034 for chromosome I and CP011035 for chromosome II. The sequence associated contextual (meta) data are MIxS compliant (Table 1) (Yilmaz et al., 2011).

Acknowledgments

The work was supported by the National Science Foundation of China (grants 31290230, 31290231, 31470001 and 91328208) and the Fundamental Research Funds of Shandong University (2014QY006, 2015JC001).

References

- Bavishi, A., Abhishek, A., Lin, L., and Choudhary, M. (2010). Complex prokaryotic genome structure: rapid evolution of chromosome II. *Genome* **53**, 675-87.
- Bowman, J. P. (2007). Bioactive compound synthetic capacity and ecological significance of marine bacterial genus *pseudoalteromonas*. *Mar Drugs* **5**, 220-41.
- Delcher, A. L., Bratke, K. A., Powers, E. C., and Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673-9.
- Egan, E. S., Fogel, M. A., and Waldor, M. K. (2005). Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes. *Mol Microbiol* **56**, 1129-38.
- Galardini, M., Pini, F., Bazzicalupo, M., Biondi, E. G., and Mengoni, A. (2013). Replicon-dependent bacterial genome evolution: the case of *Sinorhizobium meliloti*. *Genome Biol Evol* **5**, 542-58.
- Gauthier, G., Gauthier, M., and Christen, R. (1995). 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* **45**, 755-61.
- Ivanova, E. P., Sawabe, T., Lysenko, A. M., Gorshkova, N. M., Hayashi, K., Zhukova, N. V., Nicolau, D. V., Christen, R., and Mikhailov, V. V. (2002). *Pseudoalteromonas translucida* sp. nov. and *Pseudoalteromonas paragorgicola* sp. nov., and emended description of the genus. *Int J Syst Evol Microbiol* **52**, 1759-66.
- Jha, J. K., Baek, J. H., Venkova-Canova, T., and Chatteraj, D. K. (2012). Chromosome dynamics in multichromosome bacteria. *Biochim Biophys Acta* **1819**, 826-9.
- Lagesen, K., Hallin, P., Rodland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**, 3100-8.
- Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J., and Wang, J. (2010). De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res* **20**, 265-72.
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955-64.
- Medigue, C., Krin, E., Pascal, G., Barbe, V., Bernsel, A., Bertin, P. N., Cheung, F., Cruveiller, S., D'Amico, S., Duilio, A., Fang, G., Feller, G., Ho, C., Mangenot, S., Marino, G., Nilsson, J., Parrilli, E., Rocha, E. P., Rouy, Z., Sekowska, A., Tutino, M. L., Vallenet, D., von Heijne, G., and Danchin, A. (2005). Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res* **15**, 1325-35.

- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* **35**, W182-5.
- Ochman, H. (2002). Bacterial evolution: chromosome arithmetic and geometry. *Curr Biol* **12**, R427-8.
- Prozorov, A. A. (2008). [Additional chromosomes in bacteria: properties and origin]. *Mikrobiologiya* **77**, 437-47.
- Qin, Q. L., Li, Y., Zhang, Y. J., Zhou, Z. M., Zhang, W. X., Chen, X. L., Zhang, X. Y., Zhou, B. C., Wang, L., and Zhang, Y. Z. (2011). Comparative genomics reveals a deep-sea sediment-adapted life style of *Pseudoalteromonas* sp. SM9913. *ISME J* **5**, 274-84.
- Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D., and Koonin, E. V. (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29**, 22-8.
- Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J. R., Amaral-Zettler, L., Gilbert, J. A., Karsch-Mizrachi, I., Johnston, A., Cochrane, G., Vaughan, R., Hunter, C., Park, J., Morrison, N., Rocca-Serra, P., Sterk, P., Arumugam, M., Bailey, M., Baumgartner, L., Birren, B. W., Blaser, M. J., Bonazzi, V., Booth, T., Bork, P., Bushman, F. D., Buttigieg, P. L., Chain, P. S., Charlson, E., Costello, E. K., Huot-Creasy, H., Dawyndt, P., DeSantis, T., Fierer, N., Fuhrman, J. A., Gallery, R. E., Gevers, D., Gibbs, R. A., San Gil, I., Gonzalez, A., Gordon, J. I., Guralnick, R., Hankeln, W., Highlander, S., Hugenholtz, P., Jansson, J., Kau, A. L., Kelley, S. T., Kennedy, J., Knights, D., Koren, O., Kuczynski, J., Kyrpides, N., Larsen, R., Lauber, C. L., Legg, T., Ley, R. E., Lozupone, C. A., Ludwig, W., Lyons, D., Maguire, E., Methe, B. A., Meyer, F., Muegge, B., Nakielnny, S., Nelson, K. E., Nemergut, D., Neufeld, J. D., Newbold, L. K., Oliver, A. E., Pace, N. R., Palanisamy, G., Peplies, J., Petrosino, J., Proctor, L., Pruesse, E., Quast, C., Raes, J., Ratnasingham, S., Ravel, J., Relman, D. A., Assunta-Sansone, S., Schloss, P. D., Schriml, L., Sinha, R., Smith, M. I., Sodergren, E., Spo, A., Stombaugh, J., Tiedje, J. M., Ward, D. V., Weinstock, G. M., Wendel, D., White, O., Whiteley, A., Wilke, A., Wortman, J. R., Yatsunenko, T., and Glockner, F. O. (2011). Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIXS) specifications. *Nat Biotechnol* **29**, 415-20.