High quality genome sequence and description of *Enterobacter mori* strain 5–4, isolated from a mixture of formation water and crude-oil

Fan Zhang1, Sanbao Su2†, Gaoming Yu3, Beiwen Zheng4*, Fuchang Shu2, Zhengliang Wang2, Tingsheng Xiang2, Hao Dong5, Zhongzhi Zhang5, DuJie Hou1 and Yuehui She2*

Abstract

*Enterobacter mori* strain 5–4 is a Gram-negative, motile, rod shaped, and facultatively anaerobic bacterium, which was isolated from a mixture of formation water (also known as oil-reservior water) and crude-oil in Karamay oilfield, China. To date, there is only one *E. mori* genome has been sequenced and very little knowledge about the mechanism of *E. mori* adapted to the petroleum reservoir. Here, we report the second *E. mori* genome sequence and annotation, together with the description of features for this organism. The 4,621,281 bp assembly genome exhibits a G + C content of 56.24% and contains 4,317 protein-coding and 65 RNA genes, including 5 rRNA genes.

Keywords: *Enterobacter mori* strain 5–4, Formation water, Hydrocarbon degradation, Genome

Introduction

The genus *Enterobacter* was created by Hormaeche and Edwards in 1960 [1]. Members of the genus were isolated mostly from the environment, in particular from plants and recognized as notorious plant pathogens, but were also frequently isolated from hospitals, notably in healthcare associated infections and recognized as opportunistic pathogens [2,3]. Twenty-nine validly published species and 2 subspecies have previously been recorded in the genus *Enterobacter*. However, 17 of the validly named species have been subsequently reclassified as members of 11 other genera. As of Oct 2014, this genus contains only 10 species and two subspecies [4]. As of Oct, 2014, a total of 116 *Enterobacter* strains have been sequenced and 29 genome sequences were published [5-12], however, only one genome of *E. mori* isolated from diseased mulberry roots has been sequenced [13]. *E. mori* strain 5–4 is a Gram-negative, motile, rod shaped, and facultatively anaerobic bacterium, isolated from a crude-oil well. It is worthy of note that *E. mori* strain 5–4 is capable of degrading petroleum (Additional file 1). In order to elucidate comprehensive alkane degradation pathways and adaption mechanism in *E. mori* strain 5–4, whole-genome sequence analysis was thus conducted. Here, we present a summary classification and a set of features for *E. mori* strain 5–4, together with the description of the genomic sequencing and annotation.

Classification and features

A formation water sample was collected from Karamay Oilfield, Xinjiang, China, in 2012. The water sample was preserved at −80°C immediately after collection and sent to the lab. *E. mori* strain 5–4 was isolated after cultivation on LB agar medium at 37°C. The optimum temperature for growth is 35°C, with a temperature range of 4-45°C (Table 1). Growth occurs under aerobic condition. Grows at pH 5.5-10.0, and optimally at pH 7.0. Cell morphology was examined by using scanning electron microscopy (Quanta 200, FEI Co., USA). Colonies are light yellow, smooth, circular with entire margins, with a diameter ranging 0.3-0.8 μm, and from 0.6 to 1.8 μm long (Figure 1). Themethyl red test is negative. H2S and indole are not produced. Casein and starch are not hydrolysed; gelatin is hydrolysed. Sorbitol, glycerol, tetradeccane and hexadecane are utilized as the carbon source, while lactose, rhamnose, glucose, maltose, cellobiose, galactose, raffinose...
and sucrose are not utilized. Nitrite sodium and ammonium chloride are utilized, while nitrate sodium is not reduced. Antimicrobial susceptibility test showed that this strain is susceptible to ampicillin, tetracycline, erythromycin and gentamicin, and resistant to kanamycin.

A comparative taxonomic analysis was conducted based on the 16S rRNA nucleotide sequence. The representative 16S rRNA nucleotide sequence of *Enterobacter mori* strain 5–4 was compared against the most recent release of the EzTaxon-e database [26]. CLUSTAL W was used to generate alignments with comparative sequences collected from EzTaxon-e database [27]. The alignments were trimmed and converted to the MEGA 6.06 format before phylogenetic analysis. Phylogenetic inferences were made using Neighbor-joining method based on Tamura-Nei model within the MEGA 6.06 [28]. Phylogenetic tree indicated the taxonomic status of strain 5–2, clearly classified into the same branch with species *E. mori* type strain LMG 25706T (Figure 2).

**Genome sequencing information**

**Genome project history**

*E. mori* strain 5–4 was selected for whole genome sequencing on the consideration of its potential relevance to microbial enhanced oil recovery (MEOR). The genome project is deposited in the Genome On Line Database and the draft genome sequence is deposited in GenBank under the accession JFW00000000 and consists of 36 contigs. A summary of the project information and its
association with MIGS version 2.0 compliance are shown in Table 2 [14].

**Growth conditions and DNA isolation**

*E. mori* strain 5–4 was grown aerobically in Luria-Bertani Broth. Cells in late-log-phase growth were harvested and lysed by EDTA, lysozyme, and detergent treatment, followed by proteinase K and RNase digestion. Genomic DNA was extracted using the DNeasy blood and tissue kit (Qiagen, Germany), according to the manufacturer’s recommended protocol. The quantity of DNA was measured by the NanoDrop Spectrophotometer and Cubit. Then 10 μg of DNA was sent to BGI (Shenzhen, China) for sequencing on a Hiseq2000 (Illumina, CA) sequencer.

**Genome sequencing and assembly**

Genomic DNA sequencing of *E. mori* strain 5–4 was performed using Solexa paired-end sequencing technology (HiSeq2000 system, Illumina). One DNA library was generated (450 bp insert size, with Illumina adapter at both end, detected by Agilent DNA analyzer 2100), then sequencing was performed with a 2 x 100 bp pair end sequencing strategy. Finally, a total of 6,652.30 M bp data was produced and quality control was performed with the following criteria: 1) Reads linked to adapters at both end were considered as sequencing artifacts then removed. 2) Bases with quality index lower than Q20 at both end was trimmed. 3) Reads with ambiguous bases (N) were removed. 4) Single qualified reads were discarded (In this situation, one read is qualified but its mate is not). Filtered 687.39 M clean reads were assembled into scaffolds using the Velvet version 1.2.07 with parameters “-scaffolds no” [29], then we use a PAGIT flow [30] to prolong the initial contigs and correct sequencing errors to arrive at a set of improved scaffolds.

**Genome annotation**

Predict genes were identified using Glimmer version 3.0 [31], tRNAscan-SE version 1.21 [32] was used to find tRNA genes, whereas ribosomal RNAs were found by using Rfam version 1.2 [33]. To annotate predict genes, we used HMMPR version 3.0 [34] to align genes against Pfam version 27.0 [35] (only pfam-A was used) to find genes with conserved domains. KAAS server [36] was used to assign translated amino acids into KEGG Orthology [37] with SBH (single-directional best hit) method. Translated genes were aligned with COG database [38,39] using NCBI blastp (hits should have scores no less than 60, e value is no more than 1e-6). To find genes with hypothetical or putative function, we aligned genes against NCBI nucleotide sequence database database (nt database was downloaded at Sep 20, 2013 ) by using NCBI blastn, only if hits have identity no less than 0.95, coverage no less than 0.9 , and reference gene had annotation of putative or hypothetical. To define genes with signal peptide, we use SignaIP version 4.1 [40] to identify genes with signal peptide with default parameters. TMHMM 2.0 [41] was used to identify genes with transmembrane helices.

**Genome properties**

The draft genome sequence of *E. mori* strain 5–4 was assembled into 36 scaffolds with a assembly genome size of 4,621,281 bp and a G + C content of 56.2% (N50 is 358,174 bp). These scaffolds contain 4317 coding sequences (CDSs), 60 tRNAs (excluding 0 Pseudo tRNAs) and incomplete rRNA operons (3 small subunit rRNA and 2 large subunit rRNAs). A total of 980 protein-coding genes were assigned as putative function or hypothetical proteins. 3625 genes were categorized into COGs functional...
groups (including putative or hypothetical genes). The properties and the statistics of the genome are summarized in Table 3 and Table 4.

**Genome comparison**

Genome alignment between *E. mori* 5–4 (JFHW00000000) and *E. mori* type strain LMG 25706 T (AEXB00000000) was performed by using Mauve [42]. Orthology identification was carried out by a modified method introduced by Lerat [43]. Genome alignment showed that some functional regions are highly homologous between these two assemblies. The alignment also reveals some discrepancies between them, some short stretches of LMG 25706 T genome absent from the contigs in 5–4 (Figure 3A). However, two alkane 1-monooxygenase, one alkanesulfonate monooxygenase, one putative alkanesulfonate transporter, one putative sulfate permease and one alkanesulfonate transporter permease subunit were identified in the genome. Alkane 1-monooxygenase was found as one of the key enzymes responsible for the aerobic transformation of n-alkanes [44].

### Table 3 Genome statistics

| Attribute                          | Value         | % of total |
|------------------------------------|---------------|------------|
| Genome size (bp)                   | 4,621,281     | 100.00     |
| DNA Coding region (bp)             | 4,117,467     | 89.10      |
| DNA G + C content (bp)             | 2,599,117     | 56.24      |
| DNA scaffolds                       | 36            |            |
| Total genes                        | 4,322         | 100.00     |
| Protein-coding genes               | 4,317         | 99.88      |
| RNA genes                          | 65            | 1.51       |
| Pseudo genes                       | 17            | 0.39       |
| Genes with function prediction     | 980           | 22.67      |
| Genes assigned to COGs             | 3,625         | 83.87      |
| Genes assigned to Pfam domains     | 3,995         | 92.43      |
| Genes with signal peptides         | 420           | 9.72       |
| Genes with transmembrane helices   | 1,085         | 25.10      |

The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome.

### Table 4 Number of genes associated with the general COG functional categories

| Code | Value | % age | Description                                                                 |
|------|-------|-------|-----------------------------------------------------------------------------|
| J    | 202   | 4.68  | Translation, ribosomal structure and biogenesis                            |
| A    | 1     | 0.02  | RNA processing and modification                                             |
| K    | 400   | 9.27  | Transcription                                                               |
| L    | 149   | 3.45  | Replication, recombination and repair                                        |
| B    | 1     | 0.02  | Chromatin structure and dynamics                                            |
| D    | 59    | 1.37  | Cell cycle control, mitosis and meiosis                                      |
| V    | 146   | 3.38  | Defense mechanisms                                                          |
| T    | 228   | 5.28  | Signal transduction mechanisms                                              |
| M    | 266   | 6.16  | Cell wall/membrane biogenesis                                               |
| N    | 136   | 3.15  | Cell motility                                                               |
| U    | 130   | 3.01  | Intracellular trafficking and secretion                                      |
| O    | 176   | 4.08  | Posttranslational modification, protein turnover, chaperones                 |
| C    | 295   | 6.83  | Energy production and conversion                                            |
| G    | 499   | 11.56 | Carbohydrate transport and metabolism                                       |
| E    | 604   | 13.99 | Amino acid transport and metabolism                                         |
| F    | 94    | 2.18  | Nucleotide transport and metabolism                                         |
| H    | 230   | 5.33  | Coenzyme transport and metabolism                                           |
| I    | 120   | 2.78  | Lipid transport and metabolism                                              |
| P    | 421   | 9.75  | Inorganic ion transport and metabolism                                       |
| Q    | 134   | 3.10  | Secondary metabolites biosynthesis, transport and catabolism                 |
| R    | 720   | 16.68 | General function prediction only                                             |
| S    | 361   | 8.36  | Function unknown                                                            |
| -    | 333   | 7.71  | Not in COGs                                                                 |

The total is based on the total number of protein coding genes in the annotated genome.
alkanesulfonate monooxygenase and alkanesulfonate transporter may be responsible for organosulfur compound degradation [45]. Comparison of these two strains revealed the presence of a large core-genome (Figure 3B). They shared 3555 CDS in the genome. In addition, 759 CDS from the 5–4 genome were classified as unique, while 1097 CDS from the LMG 25706 T genome were classified as unique. Our genomic data will provide an excellent platform for further improvement of this organism for potential application in bioremediation.
Conclusions
Here, we report the second draft genome sequence and description of *E. mori*, which was isolated from a mixture of formation water and crude-oil. The genome revealed two alkaline 1-monooxygenase, one alkanealcohol mono-oxygenase, one putative alkanealcohol transporter, one putative sulfate permease and one alkanealcohol transporter permease subunit. Our genomic data of strain 5-4 provide a vast pool of genes involved in hydrocarbon degradation and an excellent platform for further improvement of this organism for potential application in bioremediation of oil-contaminated environments. And further comparative genomic study between stain 5-4 and other *Enterobacter* strains will give us a better understanding of the evolution of environmental bacteria towards industrial application.

Additional file

**Additional file 1: Figure S1.** Crude-oil and liquid paraffin degradation of *E. mori* 5–4. (A) Bio-degradation of crude-oil by *E. mori* 5–4 after 4-days incubation; (B) Negative control of crude-oil degradation; (C) Bio-degradation of liquid paraffin by *E. mori* 5–4 after 4-days incubation; (D) Negative control of liquid paraffin degradation.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
FZ, SBS, GMY, FCS and BWZ performed the microbiology and molecular biology studies; FZ, BWZ, HD, and ZLW performed the sequencing, annotation and genomic analysis; BWZ, YHS, ZZZ and TSX wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements
This study was supported by the National Natural Science Foundation of China (Grant No. 81301461 and No. 51474034, 863 Program (Grant No. 2013AAA064402) of the Ministry of Science and Technology, Zhejiang Provincial Natural Science Foundation of China (Grant No. LQ13H190002) and the Scientific Research Foundation of Zhejiang Provincial Health Bureau (Grant No. 2012KYB083).

Author details
1 The Key Laboratory of Marine Reservoir Evolution and Hydrocarbon Accumulation Mechanism, School of Energy Resources, China University of Geosciences, Beijing, China. 2 College of Chemistry and Environmental Engineering, Yangtze University, Jingzhou, China. 3 College of Petroleum Engineering, Yangtze University, Jingzhou, China. 4 State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, Zhejiang University, Hangzhou, China. 5 State Key Laboratory of Heavy Oil Processing, China University of Petroleum, Beijing, China.

Received: 14 May 2014 Accepted: 24 November 2014
Published: 27 February 2015

References
1. Homarovec EEP. A proposed genus Enterobacter. Int Bact Bacterial Nomen Taxon 1960; 10:71–74.
2. Zhu B, Lou MM, Xie GL, Wang GF, Zhou Q, Wang F, Fang Y, Su T, Li B, Duan YP. Enterobacter mori sp. nov., associated with bacterial wilt on Morus alba L. Int J Syst Evol Microbiol 2011; 61:2769–2774.
3. Mezzatesta ML, Gona F, Stefanis S. Enterobacter cloacae complex: clinical impact and emerging antibiotic resistance. Future Microbiol 2012; 7:887–902.
4. Taxonomic Abstract for *Enterobacter*. The NamesforLife Abstracts. Garrity GW, Parker CT, Ed. NamesforLife, LLC, 2014. http://doi.org/10.1601/tx.3148.
5. Deangels KM, D’Haeseleer PE, Chvátov D, Fortney JL, Khudyakov J, Simmons B, Woz-Harkin AP, Davenport KW, Goodlin L, Chen A, Ivanova N, Krypides NC, Mavromatis K, Woyke T, Hazen TC. Complete genome sequence of “Enterobacter lignolyticus” SGT1. Stand Genomic Sci 2011; 6:619–65.
6. Humann JL, Wildung M, Cheng CH, Lee T, Stewart IE, Drew IC, Trippett EW, Main D, Schroeder BK. Complete genome of the onion pathogen *Enterobacter cloacae* ECWSU1. Stand Genomic Sci 2011; 7:297–286.
7. Humann JL, Wildung M, Pouchnik D, Bates AA, Drew JC, Trippett UN, Trippett EW, Main D, Schroeder BK. Complete genome of the switchgrass endophyte *Enterobacter cloacae* P101. Stand Genomic Sci 2014; 9:726–734.
8. Khanna N, Ghosh AK, Huntemann M, Deshpande S, Han J, Chen A, Kyrpides N, Mavromatis K, Szeto E, Markowitz V, Ivanova N, Pagani I, Pati A, Pitluck S, Nolan M, Woyke T, Teshima H, Chentouf O, Dalgalarr K, Davenport K, Gu W, Munck C, Zhang X, Bruce D, Setter C, Xu Y, Quintana B, Retenga K, Kunde Y, Green L, et al. Complete genome sequence of *Enterobacter* sp. ITT-B08: A potential microbial strain for high rate hydrogen production. Stand Genomic Sci 2013; 8:359–369.
9. Lagier JC, El Karkouri K, Mishra AK, Robert C, Rasoul F, Poirier PE. Non-contiguous finished genome sequence and description of *Enterobacter massiliensis* sp. nov. Stand Genomic Sci 2013; 7:399–412.
10. Minogue TD, Dalgalarr K, Davenport K, Bishop-Ully KA, Bruce DC, Chan PS, Coyne SR, Chentouf O, Freitas T, Frey KG, Jaissle J, Korotkova L, Ladner JT, Palacios GF, Redden CL, Xu Y, Johnson SL. Draft Genome Assemblies of *Enterobacter* aerogenes CDC 0003–71, *Enterobacter cloacae* CDC 442–68, and *Pantoaea agglomerans* UA 0804–01. Genome Announce 2014; 20140730–14.
11. Wozel K, Gwinn-Giglio M, Nadendla S, Shefchek K, Ruppel S. Genome sequence of *Enterobacter* radiocictins DSM16656(T), a plant growth-promoting endophyte. J Bacteriol 2012; 194:5469.
12. Shin SH, Kim S, Kim JY, Lee S, Lim Y, Oh MW, Kim YR, Lee J, Yang KS. Complete genome sequence of *Enterobacter aerogenes* KCTC 2190. J Bacteriol 2012; 194:2373–2374.
13. Zhu B, Zhang GQ, Lou MM, Tian WX, Li B, Zhou XP, Wang GF, Liu H, Xie GL, Jin GL. Genome sequence of the *Enterobacter* mori type strain, LMG 25706, a pathogenic bacterium of *Morus alba* L. J Bacteriol 2011; 193:3670–3671.
14. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, Ashburner M, Axelrod N, Baldaus S, Ballard S, Boone J, Carchane G, Cole J, Dawyndt P, De Vos P, DePamphilis C, Edwards R, Faruque N, Feldman R, Gilbert J, Glina P, Gökner FQ, Goldstein P, Gurănică R, Haft D, Hancock D, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol 2008; 26:514–517.
15. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA 1990; 87:4576–4579.
16. Garrity GM, Liblum T. Phylum XIV. Proteobacteria phyl. nov. In *Bergey’s Manual of Systematic Bacteriology*, Second Edition. 2. Part B. Edited by Garrity GA, Brenner DJ, Krieg NR, Staley JT. New York: Springer; 2005:1.
17. Garrity A. Validation of publication of new names and new combinations previously effectively published outside the IJSEM. Int J Syst Evol Microbiol 2005; 55:2235–2238.
18. Garrity GM, Liblum T. Class III. Gammaproteobacteria class. nov. In *Bergey’s Manual of Systematic Bacteriology*, Second Edition, Volume 2. Edited by Brenner DJ, Krieg NR, Staley JT, Garrity GA. New York: Springer; 2005:1.
19. Garrity GM. Taxonomic Outline of the Archaea and Bacteria. In *Bergey’s Manual of Systematic Bacteriology*, Volume 1. 2nd edition. Edited by Garrity GA, Boone DR, Castenholz RW, New York: Springer; 2001:155–166.
20. Skerman VBD, Sneath PHA. Approved lists of bacterial names. Int J Syst Bacteriol 1980; 30:225–420.
21. Rahn O. New principles for the classification of bacteria. Zentralblatt für Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene. Abteilung II 1937; 96:273–286.
22. Commission. J. Conservation of the family name Enterobacteriaceae, of the name of the type genus, and designation of the type species *OPINION NO. 15*. Int Bact Bacterial Nomen Taxon 1958; 6:73–74.
23. Homarovec EEP. A proposed genus Enterobacter, Int Bact Bacterial Nomen Taxon 1960; 10:71–74.
24. Board. E. *OPINION 28* rejection of the bacterial generic name *Cloaca* Castellani and Chalmers and acceptance of *Enterobacter* Hor-maeche and Edwards as a bacterial generic name with type species *Enterobacter*
cloacae (Jordan) Hornaaeche and Edwards. Int Bull Bacte-riol Nomenc.

Taxon 1963; 13:28.

25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25:25–29.

26. Kim OS, Cho YJ, Lee K, Yoon SH, Kim M, Na H, Park SC, Jeon YS, Lee JH, Yi H, Won S, Chun J. Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. Int J Syst Evol Microbiol. 2012; 62:716–721.

27. Larkin M, Blackshields G, Brown N, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Clustal W and Clustal X version 2.0. Bioinformatics 2007; 23:2947–2948.

28. Tamura K, Stecher G, Peterson D, Filipski A, Kuma S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 2013, 30:2725–2729.

29. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008; 18:821–829.

30. Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD. Application to complete genomes. J Mol Biol 2001; 207:567–580.

31. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 2007; 23:676–679.

32. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997; 25:955–964.

33. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. Rfam: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007; 35:3100–3108.

34. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: signal peptides from transmembrane regions. J Mol Biol 2001; 306:567–580.

35. Van Hamme JD, Bottos EM, Bilbey NJ, Brewer SE. Genomic and proteomic characterization of Gordonia sp. NB4–1Y in relation to 6: 2 fluorotelomer sulfonate biodegradation. Microbiology 2013; 159:1618–1628.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Cite this article as: Zhang et al: High quality genome sequence and description of Enterobacteriaceae strain 5–4, isolated from a mixture of formation water and crude-oil. Standards in Genomic Sciences 2015 10:9.