First report of the occurrence and whole-genome characterization of *Edwardsiella tarda* in the false killer whale (*Pseudorca crassidens*)

Running title: *EDWARDSIELLA TARDA FROM THE CETACEAN*

Kyunglee LEE¹,†, Hye Kwon KIM², Sung-Kyun PARK², Hawsun SOHN¹, Yuna CHO¹, Young-Min CHOI¹, Dae Gwin JEONG², Ji Hyung KIM²,*

¹ Cetacean Research Institute (CRI), National Institute of Fisheries Science (NIFS), Ulsan 44780, Republic of Korea

² Infectious Disease Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 34141, Republic of Korea

* Corresponding author. Ji Hyung Kim, Ph.D.

Mailing address: Infectious Disease Research Center, Korea Research Institute of Bioscience and Biotechnology, 125 Gwahangno, Yuseong-gu, Daejeon 34141, Republic of Korea

Phone: (82) 42 879 8272, Fax: (82) 42 879 8498, E-mail: kzh81@kribb.re.kr
Abstract

Although several *Edwardsiella tarda* infections have been reported, its pathogenic role in marine mammals has not been investigated at the genome level. We investigated the genome of *E. tarda* strain KC-Pc-HB1, isolated from the false killer whale (*Pseudorca crassidens*) found bycaught in South Korea. The obtained genome was similar to that of human pathogenic *E. tarda* strains, but distinct from other *Edwardsiella* species. Although type III and VI secretion systems, which are essential for the virulence of other *Edwardsiella* species, were absent, several virulence-related genes involved in the pathogenesis of *E. tarda* were found in the genome. These results provide important insights into the *E. tarda* infecting marine mammals and give valuable information on potential virulence factors in this pathogen.

**Keywords:** *Edwardsiella tarda*, marine mammal, pathogen, virulence factor
The genus *Edwardsiella*, which is a member of the family Enterobacteriaceae (Proteobacteria: Gammaproteobacteria), comprises five valid species, namely, *E. anguillarum*, *E. hoshinae*, *E. ictaluri*, *E. piscicida*, and *E. tarda* [26]. Among those species, *E. tarda* is considered a pathogenic inhabitant of animals including fish, reptiles, amphibians, and birds and is associated with opportunistic zoonotic infections in humans [1]. However, the classification of *E. tarda* has been a source of controversy, and therefore, several phenotypic and genetic analyses, including whole genome sequencing, have been conducted to understand the diversity and pathogenicity of this bacterium [5, 14, 23, 24, 26]. These studies demonstrated that isolates historically classified as *E. tarda* actually represent three genetically distinct taxa with various degrees of pathogenicity in different hosts, and almost all the pathogenic fish isolates were re-assigned as *E. anguillarum* and *E. piscicida* [5, 24]. Nevertheless, bacteria currently defined as *E. tarda* still contain several fish pathogenic isolates originating from disease outbreaks in catfish aquaculture in the 1970s and 1980s [24], as well as the human pathogenic strain ATCC 23685 [23], thus suggesting that *bona fide* *E. tarda* might have zoonotic potential.

Although several pathogens were recently recognized as causative agents of emerging infectious diseases in marine mammals [31, 32], little information is currently available on bacterial infections that might pose human health risks. Among those, *E. tarda* has been considered an opportunistic pathogen presumed to cause illness or death in the sperm whale (*Physeter macrocephalus*) [9], killer whale (*Orcinus orca*) [13], beluga whale (*Delphinapterus leucas*) [17], harbor porpoise (*Phocena phocena*) [8], and pinnipeds [21]. However, its pathogenic role in wild marine mammals remains unclear owing to some bacterial findings in free-ranging bottlenose dolphins (*Tursiops truncatus*) [4, 28] and the limitations of genetic (or genomic) information on bacteria isolated from marine mammals.

Since 2016, we have investigated the potential pathogens that can colonize and establish
infection in endangered marine mammals present in coastal waters in the Republic of Korea [20]. In this study, we present the complete genome of *E. tarda* strain KC-Pc-HB1, which was isolated from a false killer whale (*Pseudorca crassidens*) found bycaught in 2017 along the South Sea (Republic of Korea). We aimed to provide genomic insights into the *E. tarda* infecting marine mammal species, and obtain useful information for the evaluation of its potential pathogenicity in those endangered species.

The general features and MIXS mandatory information for *E. tarda* strain KC-Pc-HB1 are summarized in Table 1. The bacterial strain was originally isolated from the blood collected from the heart of an adult female false killer whale (>480 cm in length, voucher no. CRI007391) found bycaught from troll fisheries in March 2017 along the South Sea (34°13'12.0"N 128°21'00.0"E, Republic of Korea). The motile, gram-negative, and flagellated straight-rod isolate (designated KC-Pc-HB1) was oxidase-negative and catalase-positive, and showed β-hemolysis on 5% sheep blood agar (Hanil Komed, Seongnam, Republic of Korea) after 24 hr of incubation at 37°C. The 16S rRNA of the isolate (MF973094) showed 99.9% similarity with the type strain of *E. tarda* ATCC 15947 (NR_024770) and *E. hoshinae* ATCC 33379 (AB682272) in the GenBank database, respectively. Sterile swabs from the blowhole and anus of the carcass were also collected and cultured under the same conditions mentioned above, and the same bacterium, which possessed 100% identical 16S rRNA sequence to that of the blood-isolate KC-Pc-HB1, was obtained from the both samples. Because the 16S rRNA was not able to discriminate the isolate as the species level, sodB sequence in KC-Pc-HB1 were obtained and used for the phylogenetic analysis, according to the previous report [24]. The resultant maximum-likelihood phylogeny indicated that the isolate KC-Pc-HB1 was well clustered with *bona fide* *E. tarda* strains (Fig. 1). Based on these results, KC-Pc-HB1 was classified as the species *tarda* and finally designated as *E. tarda* strain KC-Pc-HB1.
Genomic DNA of *E. tarda* KC-Pc-HB1 was obtained using a DNeasy blood and tissue kit (Qiagen Korea Ltd., Seoul, Republic of Korea), and it was sequenced at Macrogen Inc. (Seoul, Republic of Korea) according to the method reported by Lee *et al.* [20], using a hybrid approach with a PacBio RS II system (Pacific Biosciences, Menlo Park, CA, U.S.A.) and HiSeq 2000 instrument (Illumina, San Diego, CA, U.S.A.). The PacBio long read data (997,959,985 bp, 128,134 reads) were *de novo* assembled by the Hierarchical Genome Assembly Process program (ver. 3.0), and the Illumina pair end reads (3,716,099,262 bp, 36,793,062 reads) were mapped to the assembled contigs to improve the accuracy of the sequenced genome. Genome annotation was performed using the NCBI’s Prokaryotic Genome Annotation Pipeline (http://www.ncbi.nlm.nih.gov/books/NBK174280/), and functional categories of ORFs were analyzed by a PSI-BLAST search against the Clusters of Orthologous Groups (COG) database [30], with an E-value cutoff of 1E-4 and an identity cutoff of 20%. Bacterial tRNAs and rRNAs were respectively analyzed using tRNAscan-SE 1.21 [22] and RNAmmer 1.2 [18], and prophages were detected using PHASTER [2].

The sequenced *E. tarda* genome contained 3,720,168 bp consisting of one chromosome and one plasmid (designated pEh-Pc1) (**Fig. 2A** and **Table 2**). The final assembled circular chromosome of KC-Pc-HB1 was 3,638,764 bp (G+C content, 57.3%), and encoded 3,371 genes, 3,238 coding sequences (CDS), 28 rRNAs (5S, 16S, and 23S), 101 tRNAs, and four noncoding RNAs. The result of the G+C content analysis of the isolate also supported those of a previous study [14], showing differences in G+C content between the groups of factual *E. tarda* and other *Edwardsiella* species. The plasmid pEh-Pc1 was 81,404 bp (G+C content, 52.0%), and encoded 90 CDS including several genes associated with plasmid conjugation (*traC, traD, traL, traN*, and *traX*), thus revealing the genetic basis for its capability to transfer between bacteria. Additionally, six prophage regions (3 intact, 2 questionable, and 1 incomplete) and one additional incomplete prophage region were respectively detected in the
chromosome and plasmid pEh-Pc1 (Supplementary Table S1).

To assess overall genome similarity between KC-Pc-HB1 and other related Edwardsiella species, the average nucleotide identity (ANI) values were analyzed using the OrthoANI algorithm [19]. OrthoANI values were obtained, and a related phylogenetic tree was constructed based on OrthoANI analysis of the available representative bona fide E. tarda genomes (ATCC 15947T, ATCC 23685, and NCIMB 2034) in GenBank, and the four respective type strains of the other related Edwardsiella species (E. anguillarum ET080813T, E. hoshinae ATCC 33379T, E. ictaluri ATCC 33202T, and E. piscicida ET883T) using the orthologous ANI tool. The resulting phylogenetic trees based on OrthoANI values for KC-Pc-HB1 and other related strains indicated that the genome of the isolate was most similar (>99%) to E. tarda ATCC 15947T, isolated from human fecal samples [12], and showed relatively low genome similarity (≤88%) to the other four Edwardsiella species (Fig. 2B).

The COG functional category analysis of E. tarda KC-Pc-HB1 revealed that the functional genes encoded on the bacterial chromosome were mainly involved in COG categories of J (translation, ribosomal structure, and biogenesis), K (transcription), M (cell wall/membrane/envelope biogenesis), C (energy production and conversion), G (carbohydrate transport and metabolism), and E (amino acid transport and metabolism), whereas 5.5% and 7.9% of the predicted genes were involved in S (function unknown in COG database) and failed to find a match in the database, respectively. In addition, several functional genes (>37%) encoded on plasmid pEh-Pc1 did not have matches in the COG database, and the remaining genes were mainly involved in L (replication, recombination, and repair) (Supplementary Fig. S1).

Although the pathogenesis of E. tarda is relatively not well understood at present, previous research has shown that several factors may contribute to the virulent mechanisms of this bacterium, for example, the two types of hemolysins (HlyA and EthAB) [6, 15];
fimbrial proteins related to adhesive properties (FimABC) and killing factor MukF [27]; superoxide dismutase B (SodB) [16]; chondroitinase, urease, and EaeF (a putative *Edwardsiella* attenuation complex factor) [3, 10, 34]; and the twin arginine translocation (Tat) system consisting of tatABCDE [33]. Moreover, recent studies demonstrated that type III and type VI secretion systems (T3SS and T6SS), which contribute to the invasion and subversion of host cells, are essential for the virulence of *Edwardsiella* species [25, 29, 36]. Among the two reported hemolysin genes (*hlyA* and *ethAB*) in *Edwardsiella* species, *ethAB* was solely detected and two other putative hemolysin genes were also found to be encoded in KC-Pc-HB1. Moreover, several virulence-related genes homologous to *sodB*, *fimABC*, *mukF*, *tatABCDE*, and *chondroitinase* were encoded on the genome. However, the T3SS and T6SS homologs were not found in KC-Pc-HB1, as was shown in the other *bona fide* *E. tarda* strains [35].

Additionally, the presence of other potential virulence genes was identified by searching the Virulence Factor DataBase (http://www.mgc.ac.cn/VFs/); and consequently, several virulence-related genes involved in other pathogenic bacterial species belonging to the family Enterobacteriaceae were detected in the KC-Pc-HB1 genome (Supplementary Table S2). Moreover, the antimicrobial-resistance genes in KC-Pc-HB1 were manually searched using the ARG-ANNOT database (http://en.mediterrane-infection.com/article.php?laref=283&titre=arg-annot-). The genome was found to possess a total of three genes involving β-lactam resistance, which can also be found in other bacterial genomes belonging to the family Enterobacteriaceae, including *Edwardsiella* strains in the GenBank database, whereas no antimicrobial-resistance gene was found in the plasmid pEh-Pc1 (Supplementary Table S3). The antimicrobial-resistance of KC-Pc-HB1 was quantitatively tested according to the guidelines of the Clinical and Laboratory Standards Institute [7]; however, no acceptable phenotypical resistance was observed in all the tested
antibiotic classes (Data not shown).

According to Dunn et al. [11], *E. tarda* has been reported as one of the main causes of bacteremia and fatal septicemias in captive marine mammals; however, the evidence of its pathogenicity in wild marine mammals inevitably remains circumstantial owing to the limitations of postmortem analyses of stranded individuals and lack of genetic information of the bacterial isolates. Nevertheless, recent comparative studies indicate that *Edwardsiella* strains obtained from fish and humans are divergent [14], and most of the remaining * bona fide* *E. tarda* strains were potential human pathogenic ones [23]. Moreover, the genome of *E. tarda* KC-Pc-HB1 was almost identical to those of the * bona fide* *E. tarda* strains (Fig. 1) and possessed several virulence-related genes, making it a potentially virulent strain. Therefore, the chance for *E. tarda* transmission in humans is likely to happen during post-mortem examinations or inadvertent ingestion of infected (or contaminated) wild marine mammals and *vice versa*, because of its zoonotic potential. Consequently, more genetic information of *E. tarda* isolated from wild marine mammals is required to evaluate and clarify its potential pathogenesis in those animals. To the best of our knowledge, this is the first report of the isolation of *E. tarda* from the false killer whale and the first complete genome report of *E. tarda* found in marine mammals. The genomic data of KC-Pc-HB1 provide important insights into the biodiversity of *E. tarda* and give valuable information on potential virulence factors and antibiotic resistance for improving control strategies against this potential marine pathogen.

*Edwardsiella tarda* strain KC-Pc-HB1 was deposited in the Korean Culture Center of Microorganisms (KCCM) under KCCM 90281. The partial 16S rRNA and complete genome sequences of the strain KC-Pc-HB1 have been deposited in GenBank under accession numbers MF973094 (16S rRNA), CP023706 (chromosome), and CP023707 (plasmid pEh-Pc1).
ACKNOWLEDGMENTS

This study was supported by grants from the KRIBB Initiative program [KGM4691612] and the Global Frontier Program [2015M3A6B2063544] funded by the Ministry of Science and ICT, and a grant from the National Institute of Fisheries Science [R2017028] funded by the Ministry of Oceans and Fisheries in Republic of Korea.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References


14. Griffin, M. J., Quiniou, S. M., Cody, T., Tabuchi, M., Ware, C., Cipriano, R. C., Mauel, M. J. and Soto, E. 2013. Comparative analysis of Edwardsiella isolates from fish in the


Figure caption

**Fig. 1.** Maximum-likelihood phylogenetic tree based on the *sodB* genes from the available representative strains of *Edwardsiella* species. *Klebsiella pneumoniae* strain 342 (GenBank No. CP000964) was used as the outgroup. Numbers at the branches indicate bootstrapping values obtained with 1,000 replicates, and only bootstrap values >70% are indicated. The scale bar represents 0.05 nucleotide substitutions per site.

**Fig. 2.** (a) Circular maps of the *E. tarda* strain KC-Pc-HB1 genome. Marked characteristics are shown from the outside to the center: CDS on forward strand, CDS on reverse strand, tRNA, rRNA, GC content, and GC skew. (b) Phylogenetic trees based on OrthoANI values calculated using available genomes of bona fide *E. tarda* strains (square box) and four other related *Edwardsiella* species. The results between two strains are given in the junction point of the diagonals departing from each strain, *i.e.*, the OrthoANI value between *E. tarda* strain KC-Pc-HB1 (CP023706.1) and *E. tarda* ATCC 15947 (AFJG00000000) is 99.4%. (two-column fitting image).
**E. tarda** KC-Pc-HB1 (This study, CP023706)

- **E. tarda** ATCC 15947T (AFJG00000000)
- **E. tarda** NCIMB 2034 (MSSL00000000)
- **E. tarda** ATCC 23685 (ADGK00000000)

- **E. hoshinae** ATCC 33379T (BAUC00000000)

- **E. piscicida** ET883T (JRGQ00000000)

- **E. anguillarum** ET080813T (CP006664)

- **E. ictaluri** ATCC 33202T (AFJ100000000)

- **Klebsiella pneumoniae** 342 (CP000964)

Fig. 1.
Fig. 2.

(A) Chromosome (3,638,764 bp)

pEh-Pc1 (81,404 bp)