MICROBIAL GENOMICS

RESEARCH ARTICLE

Tekedar *et al.*, *Microbial Genomics* 2020;6

DOI 10.1099/mgen.0.000322

Comparative genomics of the fish pathogens *Edwardsiella ictaluri* 93-146 and *Edwardsiella piscicida* C07-087

Hasan C. Tekedar[1], Jochen Blom[2], Safak Kalindamar[1], Seongwon Nho[3], Attila Karsi[1] and Mark L. Lawrence[1,*]

Abstract

Edwardsiella ictaluri and *Edwardsiella piscicida* are important fish pathogens affecting cultured and wild fish worldwide. To investigate the genome-level differences and similarities between catfish-adapted strains in these two species, the complete *E. ictaluri* 93-146 and *E. piscicida* C07-087 genomes were evaluated by applying comparative genomics analysis. All available complete (10) and non-complete (19) genomes from five *Edwardsiella* species were also included in a systematic analysis. Average nucleotide identity and core-genome phylogenetic tree analyses indicated that the five *Edwardsiella* species were separated from each other. Pan-/core-genome analyses for the 29 strains from the five species showed that genus *Edwardsiella* members have 9474 genes in their pan genome, while the core genome consists of 1421 genes. Orthology cluster analysis showed that *E. ictaluri* and *E. piscicida* genomes have the greatest number of shared clusters. However, *E. ictaluri* and *E. piscicida* also have unique features; for example, the *E. ictaluri* genome encodes urease enzymes and cytochrome o ubiquinol oxidase subunits, whereas *E. piscicida* genomes encode tetrathionate reductase operons, capsular polysaccharide synthesis enzymes and vibrioferrin-related genes. Additionally, we report for what is believed to be the first time that *E. ictaluri* 93-146 and three other *E. ictaluri* genomes encode a type IV secretion system (T4SS), whereas none of the *E. piscicida* genomes encode this system. Additionally, the *E. piscicida* C07-087 genome encodes two different type VI secretion systems. *E. ictaluri* genomes tend to encode more insertion elements, phage regions and genomic islands than *E. piscicida*. We speculate that the T4SS could contribute to the increased number of mobilome elements in *E. ictaluri* compared to *E. piscicida*. Two of the *E. piscicida* genomes encode full CRISPR-Cas regions, whereas none of the *E. ictaluri* genomes encode Cas proteins. Overall, comparison of the *E. ictaluri* and *E. piscicida* genomes reveals unique features and provides new insights on pathogenicity that may reflect the host adaptation of the two species.

DATA SUMMARY

All the genomes used in this study have been deposited previously in the National Center for Biotechnology Information genome database, and their GenBank accession numbers are provided in Table 1.

INTRODUCTION

The genus *Edwardsiella* is classified in the family *Enterobacteriaceae* and contains several species that are facultative intracellular pathogens [1]. Until 2013, *Edwardsiella* consisted of three species: *Edwardsiella ictaluri*, *Edwardsiella tarda* and *Edwardsiella hoshinae*. However, based on genomics analysis, *E. tarda* was divided into three species after 2013, which resulted in two new species: *Edwardsiella piscicida* [2] and *Edwardsiella anguillarum* [3]. *E. ictaluri* is a primary bacterial pathogen that was originally identified as the causative agent of enteric septicaemia of catfish in channel catfish (*Ictalurus punctatus*) in the USA [4]. It is known to infect a broad range of other freshwater fish species in North America and Asia [5–11]. *E. ictaluri* isolates from catfish, zebrafish and tilapia have distinct

Received 07 October 2018; Accepted 09 December 2019; Published 14 February 2020

Author affiliations: [1]College of Veterinary Medicine, Mississippi State University, Mississippi State, MS, USA; [2]Bioinformatics and Systems Biology, Justus-Liebig-University Giessen, 35392 Giessen, Hesse, Germany; [3]Division of Microbiology, National Center for Toxicological Research/FDA, Jefferson, AR, USA.

***Correspondence:** Mark L. Lawrence, lawrence@cvm.msstate.edu

Keywords: *Edwardsiella ictaluri*; *Edwardsiella piscicida*; comparative genomics; type IV secretion system; orthology; mobilome.

Abbreviations: ANI, average nucleotide identity; NCBI, National Center for Biotechnology Information; SRV, score ratio value; T1SS, type I secretion system; T2SS, type II secretion system; T3SS, type III secretion system; T4SS, type IV secretion system; T5SS, type V secretion system; T6SS, type VI secretion system.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Supplementary material is available with the online version of this article: https://doi.org/10.6084/m9.figshare.8956550.v1.

genotypes and plasmid profiles [12]. Comparison of USA and Vietnamese catfish isolates revealed no detectable genetic difference, but they had distinct plasmid profiles [13]. *E. piscicida* was recently identified as a new species [2]. It was previously classified as *E. tarda* based on phenotypic and biochemical tests, but genetically it is distinguishable from *E. tarda* strains [2]. *E. piscicida* has been isolated from diseased catfish, whitefish, sea bream and bass in North America and Europe [14–17].

E. ictaluri survives in pond water and sediment under variable environmental conditions [18]. It typically causes infection in water temperatures of 22–28 °C and can actively invade healthy fish through multiple routes of entry to establish either acute or chronic infection [19–21]. *E. ictaluri* can evade innate immune mechanisms [22, 23] and replicate inside professional phagocytic cells [24] despite triggering an oxidative and nitrosative response [25–27]. The type III secretion system (T3SS) and type VI secretion system (T6SS) transfer effector proteins directly into host cells to manipulate host cell function and enable intracellular replication [28, 29]. The pathogenesis of *E. piscicida* is not well known, but it is an important emerging pathogen in USA catfish aquaculture [30]. It causes gastrointestinal septicaemia in catfish similar to *E. ictaluri*, except it has a predilection for water temperatures >28 °C.

To understand the genome level differences and similarities of these two fish pathogens, our research group sequenced the *E. ictaluri* 93-146 [31] and *E. piscicida* C07-087 [32] genomes. To learn more about the potential virulence factors of the lesser-known species *E. piscicida*, and to potentially identify genomic features that define catfish-adapted strains within these species, we conducted comparative analysis of these two genomes. To place them in perspective, we conducted average nucleotide identity (ANI), pan-genome and core-genome analyses of these two genomes along with other sequenced strains in the genus *Edwardsiella*. Based on these findings, we focused on analysis and comparison of secretion systems and mobilome elements for these two pathogens. We expect that the identified unique and shared genome features of *E. ictaluri* 93-146 and *E. piscicida* C07-087 will help elucidate the phylogeny and pathogenicity of *Edwardsiella* species.

Impact Statement

Edwardsiella ictaluri and *Edwardsiella piscicida* are important fish pathogens affecting cultured and wild fish worldwide. *E. ictaluri* is a primary bacterial causative pathogen agent of enteric septicaemia of catfish in channel catfish (*Ictalurus punctatus*) in the USA. *E. piscicida* was identified as a new species in 2013. Since then, it has been reported as a fish pathogen from different geographical locations and different sources. To investigate the genome-level differences and similarities between strains in these two species, we applied comparative genomics approaches. Our analysis revealed that there is species variation in their type VI secretion systems, which could contribute to virulence and host adaptation. Moreover, there is a relatively high number of mobile elements in *E. ictaluri* that may be a result of its type IV secretion system and lack of CRISPR-Cas systems. These mobile elements suggest that the mobilomes of these species have driven genome diversification with retention of functional pathways. Overall, this comparative genomics evaluation of two closely similar pathogens has identified important differences and provides a potential explanation of species diversification.

METHODS

Bacterial strains and genomic bioinformatics

E. ictaluri 93-146 was isolated from a natural enteric septicaemia of catfish outbreak in a commercial catfish farm in Louisiana, USA, in 1993, and *E. piscicida* C07-087 was isolated from catfish with gastrointestinal septicaemia in a commercial aquaculture pond in Mississippi, USA, in 2007. The genome sequences of *E. ictaluri* 93-146 and *E. piscicida* C07-087 were reported by our group [31, 32], and all other available genomes in the genus *Edwardsiella* were obtained from the National Center for Biotechnology Information (NCBI) genome database (as of 28/07/2017) (Table 1). For all the evaluated genomes in this study, annotation and gene prediction were conducted by RAST (Rapid Annotation using Subsystem Technology) [33] and PGAP (Prokaryotic Genome Annotation Pipeline) [34]. ANI [35] was calculated based on BLASTN results as described elsewhere [36] using the JSpecies method [37] in the EDGAR platform [38].

Genus *Edwardsiella* pan-/core-genome and singleton analyses

Pan-/core-genome analysis on all 29 *Edwardsiella* genomes was performed using EDGAR 2.0. EDGAR is based on a generic orthology criterion, which in turn is based on BLAST score ratio values (SRVs). In SRVs, BLAST scores are normalized in relation to the best hit possible, which is the BLAST result of a query gene against itself. Based on the distribution of SRVs in the dataset, a cut-off is estimated as described elsewhere [38]. In the current study, orthologs were defined as genes with a reciprocal best BLAST hit and with both single hits having an SRV above 32 %, which is comparable to the 30% cut-off used by Lerat *et al.* [39]. An exponential decay function or a Heaps' power law function was used to extrapolate the development of the size of the core or pan genome. Singleton genes for *E. ictaluri* and *E. piscicida* were identified with EDGAR 2.0. In contrast to the core- and pan-genome calculations, the singleton calculation is not based on reciprocal best BLAST hits, but on a stricter definition. Only genes that show no BLAST hits with an SRV>32% against any other analysed genome are identified as singleton genes. Identified singleton elements were downloaded and searched against the microbial virulence database (MvirDB) as described below.

Table 1. *Edwardsiella* genomes used in the comparative genomic analyses

Evaluated genomes in the genus *Edwardsiella* were obtained from the NCBI genome database (as of 28/07/2017). NA, Not available.

Species	Strain	Location	Source	Genome close level/contig numbers	Size (Mb)	G+C (mol%)	Genome accession numbers	Number of plasmids	Plasmid accession numbers	Reference
E. ictaluri	RUSVM-1	Western hemisphere	Nile tilapia	Complete	3.63064	57.40	NZ_CP020466			[81]
E. ictaluri	LADL11-100	USA	Zebrafish	Contig	3.70126	57.40	PRJNA285663	2	KR869777/KR869778	[82]
E. ictaluri	LADL11-194	USA	Zebrafish	Contig	3.70614	57.40	PRJNA285852			[82]
E. ictaluri	ATCC 33202	USA	Channel catfish	Contig	3.70369	57.60	PRJNA66365			[83]
E. ictaluri	93-146	USA	Channel catfish	Complete	3.8123	57.40	NC_012779	2	NC_002497/NC_002498	[31]
E. piscicida	RSB1309	Japan	Red Sea bream	Contig	3.83865	59.30	PRJDB1727			[84]
Edwardsiella sp.	EA181011	Israel	White grouper	Complete	3.93417	59.10	NZ_CP011364			[85]
Edwardsiella sp.	LADL05-105	USA	Tilapia	Complete	4.14204	58.80	NZ_CP011516			[86]
E. anguillarum	ET080813	PR China	Japanese eel	Complete	4.32965	58.37	NZ_CP006664	2	NZ_CP006665/NZ_CP006666	[3]
E. anguillarum	ET081126R	PR China	European eel	Contig	4.23812	58.50	NZ_JACP01000006			[3]
E. anguillarum	ET070829	PR China	Marbled eel	Contig	4.28636	58.40	NZ_JABY01000001			[3]
E. anguillarum	ET080729	PR China	Marbled eel	Contig	4.15319	58.40	NZ_JABS01000002			[3]
E. tarda	ET883	Norway	European eel	Contig	3.82967	59.60	PRJNA259400			[87]
E. tarda	FL6-60	USA	Striped bass	Complete	3.7288	59.70	NC_017309	1	NC_017318	[83]
E. piscicida	ETW41	Korea	Eel pond water	Complete	3.96904	59.48	NZ_CP019440	1	NZ_CP019441	NA
E. piscicida	JF1305	Japan	Olive flounder	Contig	3.71682	59.80	PRJDB1727			[84]
E. tarda	EIB202	PR China	Turbot	Complete	3.80417	59.67	NC_013508	1	NC_013509	[88]
E. piscicida	C07-087	USA	Channel catfish	Complete	3.85704	59.60	NC_020796			[32]
E. piscicida	ACC35.1	Europe	Turbot	Contig	3.84163	59.70	PRJNA353909			[89]
E. piscicida	S11-285	USA	Channel catfish	Complete	3.92677	59.59	NZ_CP016044	1	NZ_CP016445	[90]
E. hoshinae	ATCC 33379	France	Female puffin	Contig	3.70661	56.80	PRJDB228			[91]
E. hoshinae	ATCC 35051	USA	Monitor lizard	Complete	3.81165	56.90	NZ_CP016043			[92]
E. tarda	ASE201307	PR China	Swamp eel	Contig	3.68479	57.10	PRJNA329181			NA
E. tarda	NBRC 105688	USA	Human faeces	Contig	3.61038	57.30	PRJDB227			[1]
E. tarda	ATCC 23685	USA	Human faeces	Contig	3.74457	56.80	PRJNA28661			[1]
E. tarda	ATCC 15947	USA	Human faeces	Contig	3.69469	57.10	PRJNA39897			[1]

Continued

Table 1. Continued

Species	Strain	Location	Source	Genome close level/contig numbers	Size (Mb)	G+C (mol%)	Genome accession numbers	Number of plasmids	Plasmid accession numbers	Reference
E. tarda	FL95-01	USA	Channel catfish	Complete	3.6207	57.40	NZ_CP011359			[93]
E. tarda	DT	PR China	Oscar fish	Contig	3.75958	57.00	PRJNA66369			[94]
E. tarda	NCIMB2034	USA	Unknown fish	Contig	3.76083	57.20	PRJNA357015			[95]

Genus *Edwardsiella* phylogenetic tree creation

Phylogenetic distances were calculated based on the complete core genome derived from 29 *Edwardsiella* genomes. All orthologous gene sets of the core genome were individually aligned using MUSCLE [40]. The resulting alignments were concatenated and phylogenetic distances were calculated using the neighbour-joining method in PHYLIP [41]. Tree topology was validated using 500 bootstrapping iterations.

E. ictaluri and *E. piscicida* gene comparison

Gene variation between *E. ictaluri* and *E. piscicida* was compared using comparative tools in RAST [33], which allows for potential genome-to-genome variation in annotation. This analysis identified homologous genes in *E. ictaluri* strains 93-146, LADL11-100, LADL11-194, ATCC33202 and RUSVM-1, and *E. piscicida* strains C07-087, ACC35.1, ETW41, JF1305 and S11-285. For visualization of the comparative genome analysis, BRIG (BLAST Ring Image Generator) was used [42].

E. ictaluri and *E. piscicida* orthologous gene cluster analysis

Comparisons were also conducted between *E. ictaluri* and *E. piscicida* genomes at the protein level using genome-wide analysis of orthologous clusters, which was calculated for only complete *E. ictaluri* and *E. piscicida* genome sequences using OrthoVenn [43]. With this tool, genome-wide comparisons and visualization of orthologous clusters from complete *Edwardsiella* genomes were created and visualized using the following parameters: *E* value, 1×10^{-5}; inflation value, 1.5. Protein files for each genome were downloaded from the NCBI and uploaded to OrthoVenn (http://www.bioinfogenome.net/OrthoVenn/) for identification of orthologous clusters. Protein accession numbers used can be obtained from the genome accession numbers listed in Table 1. For accuracy, only complete genomes were used in this analysis. We performed four different analyses. First, the *E. ictaluri* 93-146 genome was compared against *E. ictaluri* RUSVM-1. Second, *E. piscicida* C07-087 and other complete *E. piscicida* genomes (strains S11-285, EIB202 and FL6-60) were compared. Third, the *E. ictaluri* 93-146 genome was compared against *E. piscicida* C07-087. Finally, six *Edwardsiella* genomes were compared against each other: 2 *E. ictaluri* (strains 93-146 and RUSVM-1) and 4 *E. piscicida* (strains C07-087, S11-285, EIB202 and FL6-60).

Protein secretion systems

To detect secretion systems and their components in the evaluated *E. ictaluri* and *E. piscicida* genomes, bacterial type I, III, IV, V and VI secretion system (T1SS, T3SS, T4SS, T5SS and T6SS, respectively) proteins were identified by uploading protein sequences to MacSyDB/TXSSdb [44–46]. In addition to MacSyDB/TXSSdb, SecRet6 was used to confirm T6SS proteins [47]. A bacterial T4SS resource, SecRet4, was used to confirm identified elements from this secretion system [48]. To create the operon figures, Vector NTI was used [49].

Insertion elements

Issaga was used to identify individual insertion sequences in each *E. ictaluri* and *E. piscicida* genome [50]. After identification, results were filtered by removing false predicted elements from the final list.

Genomic islands

IslandViewer 4 was used to identify genomic islands in the *E. ictaluri* and *E. piscicida* genomes [51]. Some of the genus *Edwardsiella* members were pre-analysed in IslandViewer. For the unanalysed genomes, their GBK files were downloaded from the RAST annotation server, and *E. ictaluri* 93-146 and *E. piscicida* C07-087 genomes were used as a reference. In this tool, four different island prediction methods were used: IslandPick, SIGI-HMM, IslandPath-DIMOB and Islander. To determine potential virulence proteins encoded in genomic islands, the identified genomic islands (based on integrated results in IslandViewer 4) were downloaded and searched against MvirDB as described below.

Prophages

PHASTER (PHAge Search Tool Enhanced Release) was used to identify prophages in the *E. ictaluri* and *E. piscicida* genomes [52]. Some of the genomes were pre-computed in the PHASTER database, and some were not. For those with pre-computed genomes, nucleotide files were concatenated to serve as input files and submitted for phage element identification in the *E. ictaluri* and *E. piscicida* genomes. Depending on the result from PHAST, identified phage regions were placed into three categories: if the identified region's score was >90, these were considered intact phage regions; if it was between 70 and 90, it was considered questionable; and if it was <70, it was classified as an incomplete phage region.

Integron identification

The presence of integron elements in the *E. ictaluri* and *E. piscicida* genomes was investigated by IntegronFinder [53] based on nucleotide sequences. The threshold for clustering was 4000 bp, *attC E* value was 1, maximum value for *attC* size was 200 bp and minimum value for *attC* size was 40 bp.

CRISPR and Cas element analysis

To determine CRISPR-Cas systems and their elements in the genus *Edwardsiella*, CRISPRfinder [54] and MacSyFinder [55] were used. In MacSyFinder, maximal *E* value was set to 1.0, independent *E* value was set to 0.001 and minimal profile coverage parameters were set for Cas element identification in the genus *Edwardsiella*. Results were sorted into three categories: *mandatory*, *accessory* and *forbidden*. If the elements were ubiquitous and identifiable, they were considered *mandatory. Accessory* components could be essential but not identifiable due to rapid evolution or other reasons. If the evaluated element was partly homologous, the system identified it as a *forbidden* element.

Virulence factors

Potential virulence factors of singleton elements and genomic islands in *E. ictaluri* 93-146 and *E. piscicida* C07-087 genomes were assessed by searching them against MvirDB [56]. Local BLAST was conducted with all the predicted proteins for all evaluated genomes using CLC Genomics Workbench (version 6.5). In this analysis, BLAST results were considered significant when the *E* value was $<1\times10^{-20}$, which is a generally accepted consensus cut-off [57].

RESULTS

Genome features of the genus *Edwardsiella*

Genome summaries and features are listed in Table 1. The G+C content of *Edwardsiella* genomes ranges from 56.8 to 59.80 mol%. *E. hoshinae* genomes have the lowest G+C content, whereas *E. ictaluri* genomes have approximately 57.4 mol%, and *E. piscicida* genomes have the highest G+C content (59.7 mol%). Plasmids from *E. hoshinae* strain ET080813, four *E. piscicida* genomes (strains FL6-60, ETW41, EIB202 and S11-285) and two *E. ictaluri* genomes (strains LADL11-100 and LADL11-194) were included in the analysis.

Genus *Edwardsiella* pan-/core-genome and singleton analyses

The SRV cut-off of 32% used in the current project resulted in a mean *E* value of 1×10^{-9} and a mean identity of 84.4% for the 7 632 109 individual BLAST results that passed the SRV filter. Pan-genome analysis of the 29 *Edwardsiella* strains identified a total of 9474 genes (Fig. 1a). There were 1421 genes identified in the core genome (Fig. 1b). Using all 29 *Edwardsiella* genomes, the *E. ictaluri* 93-146 genome had 268 identified singletons, whereas the *E. piscicida* C07-087 genome had 47 singleton genes. Four of these encode *E. ictaluri* 93-146 proteins (NT01EI_2881, NT01EI_2480, NT01EI_2479, NT01EI_1391) that had significant matches in MvirDB, whereas only one of the proteins encoded by singletons in the *E. piscicida* genome (ETAC_16540) had a significant match in MvirDB (Supplementary file 1, available with the online version of this article).

ANI and phylogenetic analysis of genus *Edwardsiella*

ANI showed that all *Edwardsiella* species (*E. ictaluri*, *E. piscicida*, *E. anguillarum*, *E. hoshinae* and *E. tarda*) are distinctly different from each other (Table 2). ANI results indicated that *E. anguillarum* and *E. piscicida* are the closest related species in the genus, followed by *E. ictaluri. E. tarda* and *E. hoshinae* are more distantly related to the three other *Edwardsiella* species. Phylogenetic analysis based on the complete core genome of 29 *Edwardsiella* genomes was conducted. The tree for the 29 genomes was built out of a core of 1419 genes per genome. The core has 1227430/3682290 aa residues/bp per genome. Results showed that the five *Edwardsiella* species formed distinguishable branches (Fig. 2). Note that some of the *Edwardsiella* species labels in Fig. 2 are not accurate, because their original species designation has changed since the newer species *E.*

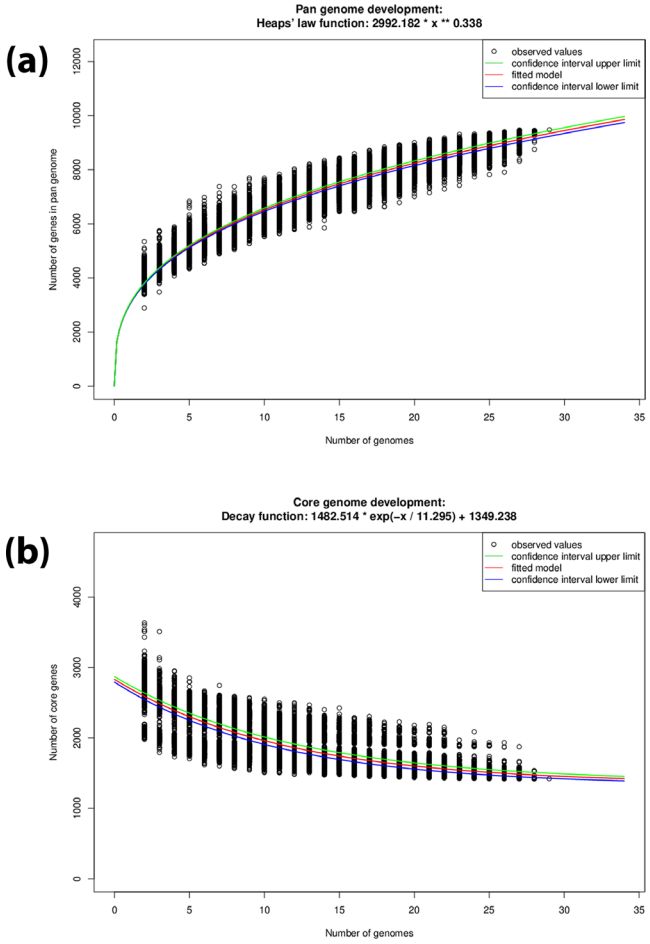


Fig. 1. Pan- versus core-genome plot analysis of genus *Edwardsiella* members. (a) Pan-genome development. (b) Core-genome development. Heaps's power law function (pan genome) was used to extrapolate the development of the size of the pan genome, and an exponential decay function (core genome) was used to display core-genome development. Extrapolation of the pan genome showed that it is open with a growth factor γ of 0.338. The predicted core-genome size is 1349 genes, which is close to the actual core genome of 1421 genes.

piscicida and *E. anguillarum* were defined; these strains are indicated with red stars.

E. ictaluri and *E. piscicida* genome structure variation

Evaluation of gene variation using all the sequenced *E. ictaluri* and *E. piscicida* genomes (complete and contig) showed that the *E. ictaluri* genomes uniquely encode urease and cytochrome o ubiquinol oxidase subunits. However, the *E. piscicida* genomes encode tetrathionate reductase, capsular polysaccharide synthesis enzymes and vibrioferrin-related proteins (Supplementary file 2). Visualization of the gene comparison showed that many of the unique genes are clustered in genomic islands and phage elements, which correlates with variation in G+C content (Fig. 3a, b).

E. ictaluri and *E. piscicida* orthology analysis

In addition to the pan-/core-genome analysis, we conducted a more focused orthology analysis of *E. ictaluri* and *E. piscicida* using four different combinations with OrthoVenn. In the first one, we performed a comparison of proteins encoded by the *E. ictaluri* 93-146 and *E. ictaluri* RUSVM-1 genomes. These two strains shared 2774 clusters of orthologous proteins. Each cluster represents a shared protein, indicating that the strains have a shared protein function. Only 13 clusters (including insertion sequences, phage, putative glycosyltransferase, 2-dihydro-3-deoxy-D-gluconate 5-dehydrogenase and NADH oxidase elements) were unique to *E. ictaluri* 93-146. Four clusters (including prophage CP4-57 regulatory protein AlpA, type I restriction enzyme *Eco*AI R and anticodon nuclease) were unique to RUSVM-1 (Fig. 4a) (Supplementary file 3).

Second, we compared the *E. piscicida* strains with complete genomes and found that the *E. piscicida* S11-285 genome encodes five unique clusters (including putative lambdoid prophage e14 repressor protein C2, ornithine decarboxylase and antitermination protein Q homologue from lambdoid prophage Qin), which is more than the *E. piscicida* C07-087, EIB202 and FL06-60 genomes. *E. piscicida* C07-087 had two unique clusters (Fig. 4b) (Supplementary file 3).

Third, we evaluated orthologous clusters encoded by the *E. ictaluri* 93-146 and *E. piscicida* C07-087 genomes and found that they share 2708 orthologous clusters. Specifically, the *E. ictaluri* 93-146 genome encodes 18 unique clusters (such as insertion elements, transposons and invasion LpaB proteins), whereas the *E. piscicida* C07-087 genome encodes 20 unique clusters (such as probable dipeptidase, platelet binding protein GspB, sn-glycerol-3-phosphate-binding periplasmic protein UgpB, transcriptional regulatory protein FixJ and cellulose synthase catalytic subunits) (Fig. 4c) (Supplementary file 3).

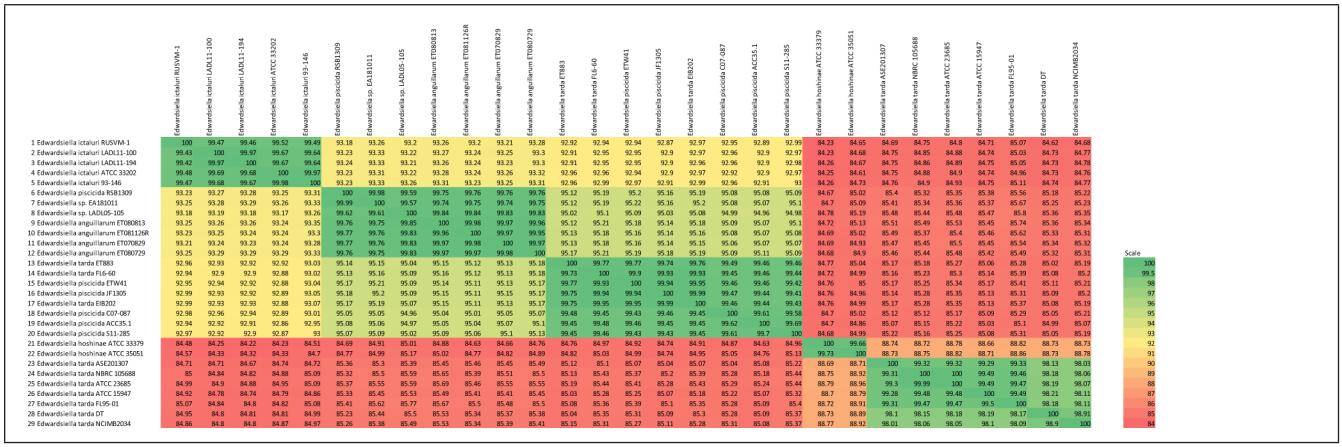
Fourth, all the *E. ictaluri* strains (93-146 and RUSVM-1) and *E. piscicida* strains (S11-285, C07-087, EIB202 and FL06-60) that were fully sequenced were evaluated. They shared 2495 clusters, and only two hypothetical proteins were uniquely encoded by the *E. piscicida* C07-087 genome, whereas the *E. ictaluri* genome encoded only two unique insertion element clusters (Fig. 4d) (Supplementary file 3).

Secretion systems

We evaluated all the secretion elements encoded in the *E. ictaluri* and *E. piscicida* genomes. All of the evaluated *E. ictaluri* and *E. piscicida* genomes encoded the T1SS, T3SS, T5SS and T6SS. T6SS-type I subtype i4b was encoded by all the *E. ictaluri* strains. However, the *E. piscicida* C07-087, S11-285 and ACC35.1 genomes encoded both T6SS-type I subtype i4b (Fig. 5a) and T6SS-type I subtype i2 (Fig. 5b). The other evaluated *E. piscicida* genomes (strains ET883, FL6-60, ETW41, JF1305 and EIB202) encoded T6SS-type I subtype i4b. Finally, *E. piscicida* strain EIB202 carried a plasmid that encodes some of the T4SS-type-T elements. In comparison, the *E. ictaluri* 93-146, LADL11-100, LADL11-194 and ATCC 33202 genomes encoded T4SS-type G (Fig. 5c; proteins listed

Table 2. ANIs of the *Edwardsiella* genomes

The strain names used in the ANI table are extracted from the NCBI genome database. ANI results indicate that *E. piscicida* RSB1309, *Edwardsiella* sp. EA181011 and LADL05-105 should be reclassified as *E. anguillarum*. *E. tarda* strains ET883, FL6-60 and EIB202 should be reclassified as *E. piscicida*.



in Table 3). Intriguingly, the *E. ictaluri* RUSVM-1 genome did not encode any of the T4SS.

Insertion sequences

Several insertion sequences were identified in the *E. ictaluri* and *E. piscicida* genomes. IS*3* family members (IS*407*, IS*51* and IS*3*) represented the only family encoded by all the evaluated *E. ictaluri* and *E. piscicida* genomes. The *E. ictaluri* 93-146, LADL11-100, ATCC 33202 and RUSVM-1 genomes encoded 11 different insertion sequence families (strain LADL11-194 lacked one insertion element from the IS*As1* family). Thus, *E. ictaluri* genomes appeared to have relatively conserved insertion sequences, while *E. piscicida* genomes showed more variable insertion sequence elements. For example, IS*1*, IS*427*, IS*110*, IS*256* and IS*L3* elements are encoded more in *E. ictaluri* genomes than *E. piscicida* genomes. IS*481* elements were unique to *E. ictaluri*, whereas IS*50*, IS*H8*, IS*6*, IS*21*, IS*91* and Tn*3* elements were unique to *E. piscicida*, but they were encoded by only some of the *E. piscicida* genomes (Table 4).

Genomic islands

The evaluated *E. ictaluri* strains had more genomic islands (RUSVM-1, 35; LADL11-100, 40; LADL11-194, 38; ATCC 33202, 39; and 93-146, 36) than *E. piscicida* strains (ET883, 34; FL6-60, 23; ETW41, 34; JF1305, 26; EIB202, 30; C07-087, 29; ACC35.1, 28; and S11-285, 36) (Fig. 3a, b). *E. ictaluri* 93-146 had 517 proteins in its genomic islands, and 212 of them had significant BLAST matches with MvirDB. The *E.*

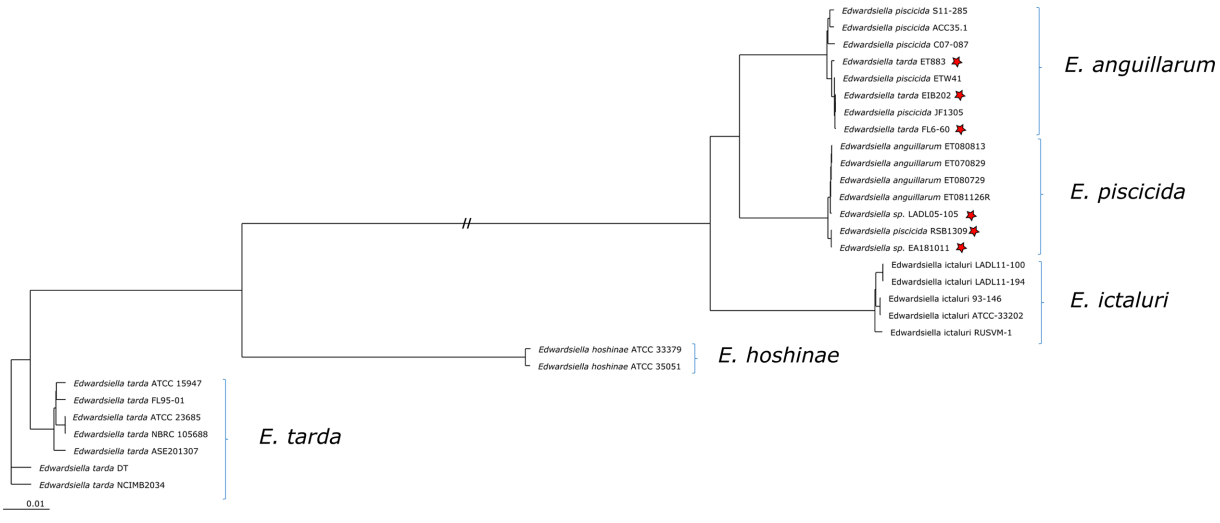


Fig. 2. Phylogenetic tree of *Edwardsiella* species based on core genomes. The main *Edwardsiella* branch was manually shortened for improved visualization ('//' indicates the shortened branch). Some of the *Edwardsiella* species labels in this figure are not accurate, because they still reflect their original species classifications; they are shown as they are currently listed in the NCBI genome database. *E. piscicida* RSB1309, *Edwardsiella* sp. LADL05-105 and *Edwardsiella* sp. EA181011 should be classified as *E anguillarum*. *E. tarda* ET883, *E. tarda* FL6-60 and *E. tarda* EIB202 should be classified as *E. piscicida*. Red stars indicate the misclassified strains.

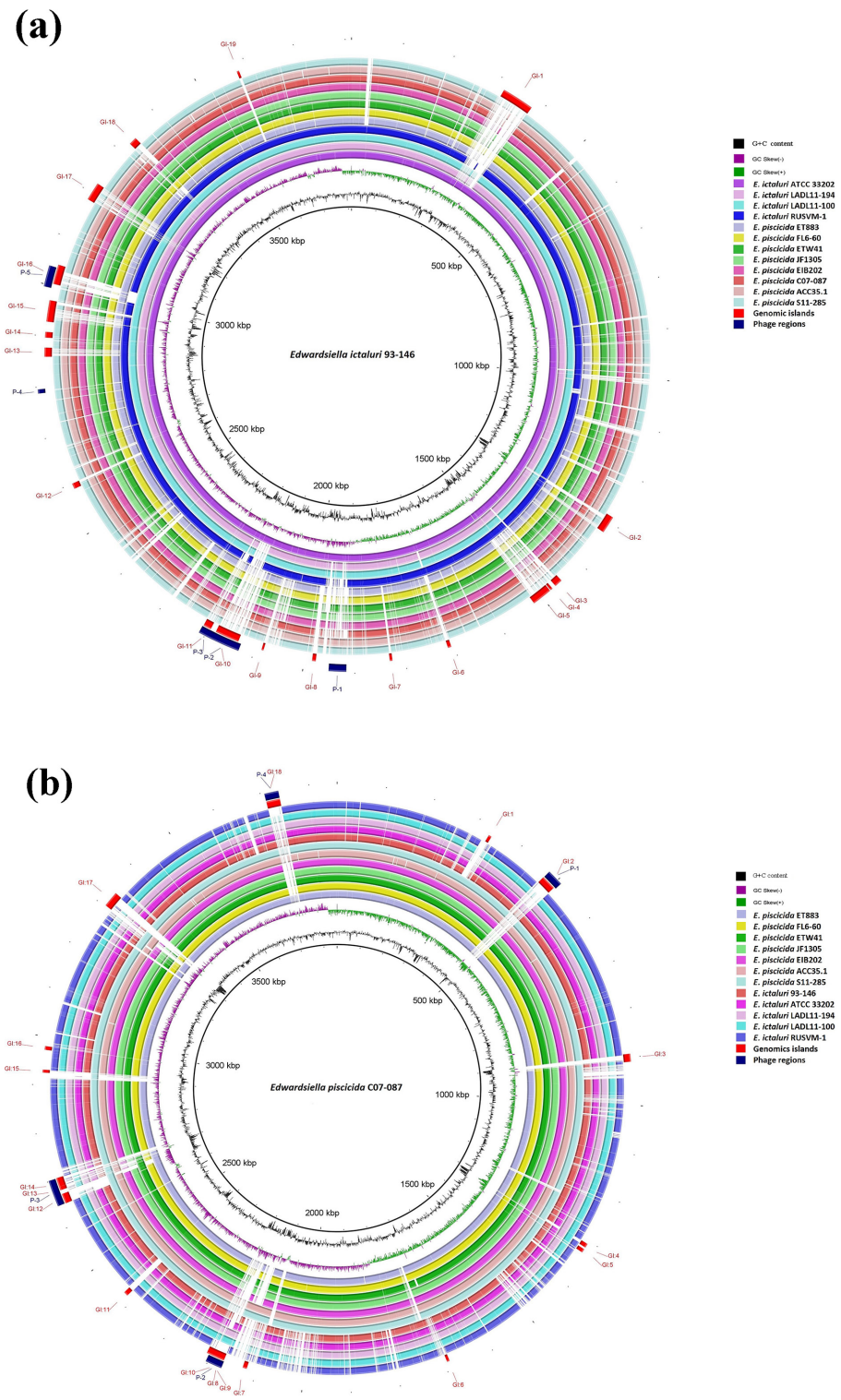


Fig. 3. Comparative circular visualization of *E. ictaluri* and *E. piscicida* genomes. The two inner rings represent the G+C content (black) and GC-skew (green/purple). The next 12 rings show gene comparisons between the reference strain and the strain listed (colour indicates the homologous gene is present). The outer two rings indicate the presence of genomic islands and phage elements in the reference strain. (a) The reference strain is *E. ictaluri* 93-146. (b) The reference strain is *E. piscicida* C07-087.

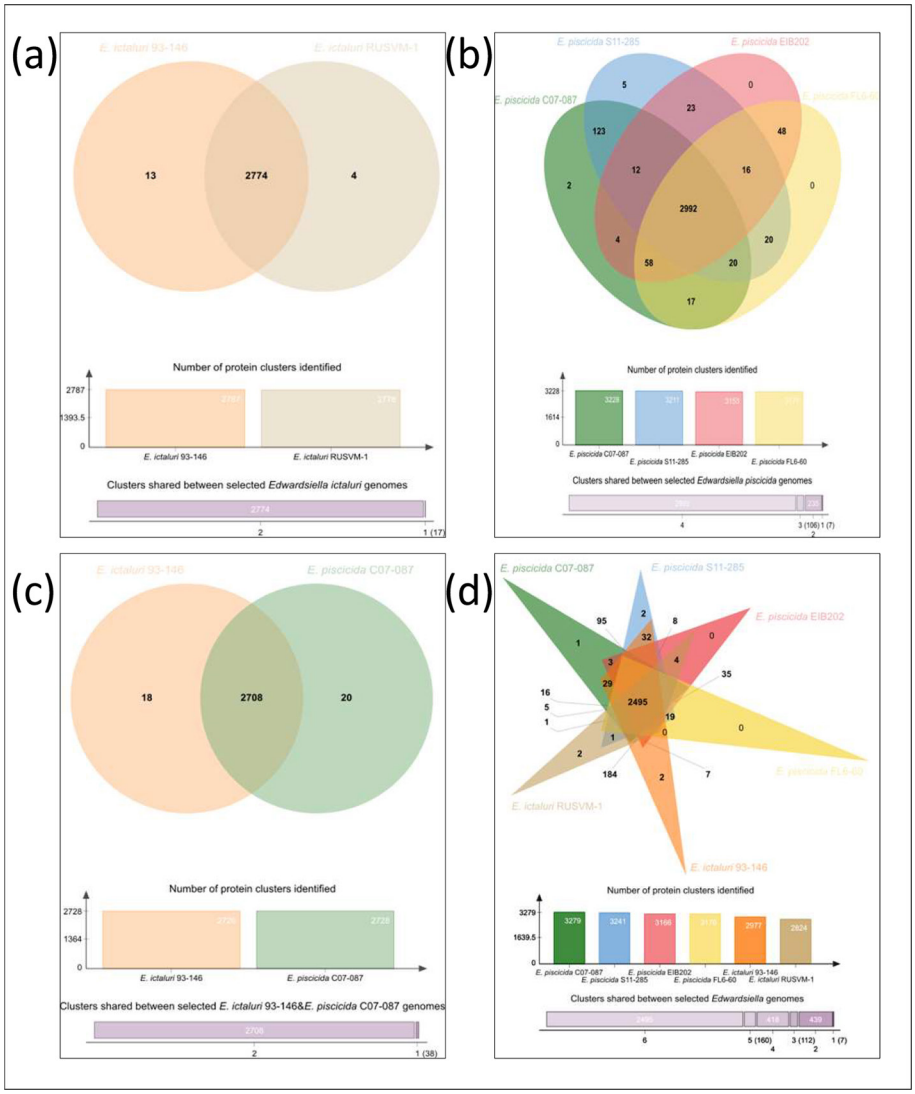


Fig. 4. Venn diagrams of protein clusters encoded by *E. piscicida* and *E. ictaluri* based on orthology. (a) *E. ictaluri* 93-146 and *E. ictaluri* RUSVM-1. (b) *E. piscicida* C07-087, S11-285, EIB202 and FL6-60. (c) *E. ictaluri* 93-146 and *E. piscicida* C07-087. (d) *E. ictaluri* strains 93-146 and RUSVM-1, and *E. piscicida* strains C07-087, S11-285, EIB202 and FL6-60. The bar charts and Venn diagrams represent the numbers of shared and unique orthologous genes of each genome.

piscicida C07-087 genome had 367 proteins in its genomic islands, and 97 of them had significant matches in MvirDB (Supplementary file 4).

Phage sequences

E. ictaluri strains tended to have more prophages than *E. piscicida* strains. *E. ictaluri* strains had a mean of 2.75 intact prophages per genome and 5.75 total phages per genome; *E. piscicida* strains had a mean of 1.5 intact prophages per genome and 3.38 total phages per genome. In detail, *E. ictaluri* strain RUSVM-1 had 5 incomplete (In); LADL11-100 had 4 intact (I) and 2 In; LADL11-194 had 4 I, 2 questionable (Q) and 1 In; ATCC 33202 had 3 I, 1 Q and 1 In; 93-146 had 3 I and 2 In. *E. piscicida* strains ET883 had 3 I, 3 Q and 1 In; FL6-60 had 1 In; pFL6-60 had 1 I; ETW41 had 2 I, 2 Q and 2 In; JF1305 had 1 Q and 1 In; EIB202 had 2 Q and 1 In; C07-087 had 4 I; ACC35.1 had 3 I and 1 Q; and S11-285 had 1 I. G+C content of *E. ictaluri* and *E. piscicida* phage regions varied between 42.25 to 56.93 mol% and 46 to 59.82 mol%, respectively. Phage elements for *E. ictaluri* 93-146 genome are shown in Fig. 3(a). *E. piscicida* C07-087 phage elements are shown in Fig. 3(b) . Details are listed in Supplementary file 5.

Integron identification

Integron analysis showed that there were no complete integrons nor any integron elements in the *E. ictaluri* genomes, while two *E. piscicida* strains (ETW41 and JF1305) encoded integron elements. Normally, complete integrons are composed of three elements: an integrase gene, an *attI* recombination site and an array of gene cassettes. The gene cassettes

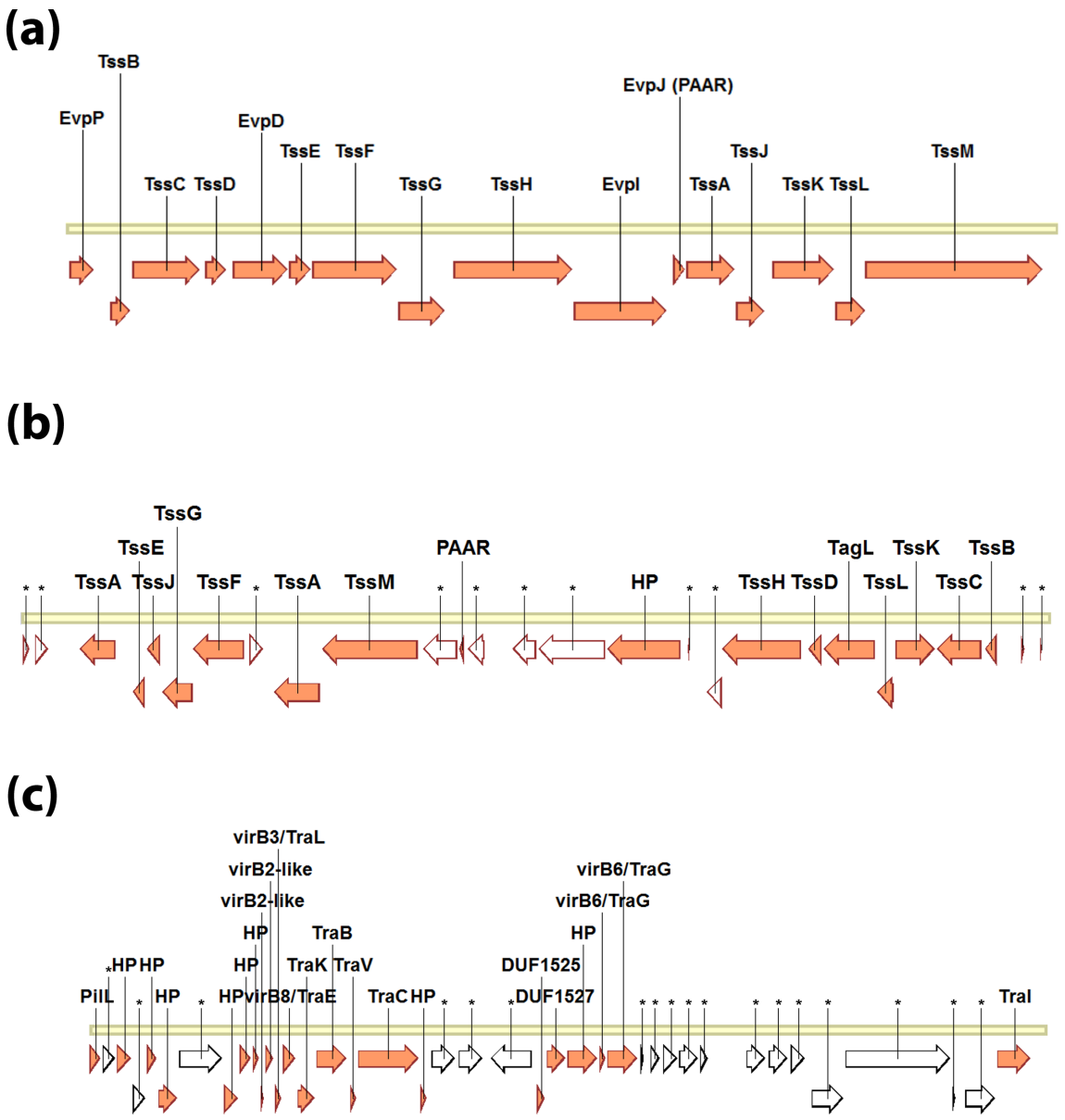


Fig. 5. Selected secretion systems in *E. ictaluri* and *E. piscicida*. (a) T6SS type I subtype i4B; (b) T6SS type I subtype i2 in *E. piscicida*; (c) T4SS in *E. ictaluri* 93-146. Orange arrows represent genes encoding elements of the secretion systems; other genes are indicated by white arrows and labelled with *.

are composed of two *attC* recombination sites. Strains ETW41 and JF1305 only encoded integrase genes and *attC* elements.

CRISPR and Cas elements

There are three major Cas systems (type I, type II and type III) with 10 different subtypes (type I-A to -F, type II-A and -B, and type III-A and -B) [55]. The *E. ictaluri* and *E. piscicida* genomes had the *csm2* gene from type III-A and the *cas3* gene from type I. Only *E. piscicida* ACC35.1 and S11-285 genomes encoded type I-E including *cas1*, *cse2*, *cas5*, *cas3*, *cas6*, *cse1*, *cas2* and *cas7* genes. Often CRISPR regions flank *cas* elements. Only *E. piscicida* ACC35.1 and S11-285 genomes had confirmed CRISPR regions, whereas the rest of the *E. piscicida* genomes encoded only questionable CRISPR regions. Furthermore, only *E. ictaluri* LADL11-100 and LADL11-194 encoded questionable CRISPR elements, and the rest of the *E. ictaluri* genomes did not carry any CRISPR elements (Supplementary file 6).

DISCUSSION

The goal of our study was to compare the available genomes of two fish pathogens, *E. ictaluri* and *E. piscicida* (including two strains that we sequenced previously, 93-146 and

Table 3. *E. ictaluri* T4SS proteins

Gene	Protein	Annotation and potential role	Availability in the *E. ictaluri* genomes				
			93-146	ATCC 33202	LADL11-194	LADL11-100	RUSVM-1
tfc2	PilL	Hypothetical protein, conjugative transfer region protein	+	+	+	+	−
tfc3	HP	Hypothetical protein, integrating conjugative element protein, PFL_4693 family	+	+	+	+	−
virB1	VirB1	Hypothetical protein, lytic transglycosylase	+	+	+	+	−
tfc5	HP	Hypothetical protein	+	+	+	+	−
t4cp2	TraD	Conjugative coupling factor TraD	+	+	+	+	−
tfc7	HP	Hypothetical protein, integrating conjugative element membrane protein, PFL_4697 family	+	+	+	+	−
tfc8	HP	Conserved hypothetical protein, RAQPRD motif	+	+	+	+	−
tfc9	VirB2-like	Hypothetical protein, integrating conjugative element protein	+	+	+	+	−
tfc10	VirB2-like	Integrating conjugative element membrane protein	+	+	+	+	−
tfc11	VirB3/TraL	Hypothetical protein	+	+	+	+	−
tfc12	VirB8/TraE	Integrating conjugative element protein	+	+	+	+	−
tfc13	TraK	Hypothetical protein	+	+	+	+	−
tfc14	TraB	Integrating conjugative element protein, bacterial conjugation TrbI-like protein	+	+	+	+	−
tfc15	TraV	Conjugative transfer region lipoprotein	+	+	+	+	−
virB4	TraC	F pilus assembly T4SS for plasmid transfer, TraC	+	+	+	+	−
tfc17	HP	Hypothetical protein, acetyltransferase	+	−	+	+	−
tfc24	DUF1525	Integrating conjugative element protein	+	+	+	+	−
tfc23	DUF1527	Integrating conjugative element protein	+	+	+	+	−
tfc22	HP	Integrating conjugative element protein	+	+	+	+	−
tfc18	VirB6/TraG	Hypothetical protein	+	+	+	+	−
tfc19	VirB6/TraG	TraG_N	+	+	+	+	−
mobB	TraI	TraI_2_C, conjugative transfer protein MobH, relaxase	+	+	+	+	+

HP, Hypothetical protein.

Table 4. Number of insertion sequence elements in the *E. ictaluri* and *E. piscicida* genomes by family

Families	Sub-groups	Typical size-range (bp)	Mechanisms	*Edwardsiella ictaluri* RUSVM-1	*Edwardsiella ictaluri* LADL11-100	*Edwardsiella ictaluri* LADL11-194	*Edwardsiella ictaluri* ATCC 33202	*Edwardsiella ictaluri* 93146	*Edwardsiella piscicida* ET883	*Edwardsiella piscicida* FL6-60	*Edwardsiella piscicida* ETW41	*Edwardsiella piscicida* JF1305	*Edwardsiella piscicida* EIB202	*Edwardsiella piscicida* C07-087	*Edwardsiella piscicida* ACC35.1	*Edwardsiella piscicida* S11-285
IS*1*		740-1180	Copy, paste, conintegrate	37	32	39	29	39						1		
IS*3*	IS*407*	1100-1400	Copy, paste	8	8	7	10	10	4	7	5	5	5	8	3	9
	IS*51*	1000-1400		5	3	4	4	4	1	6	2	2	2	2	2	8
	IS*3*	1150-1750		7	7	6	6	6	1	2	2	2	2	1	2	2
IS*481*	-	950-1300	Copy, paste	1	2	3	2	3								
IS*4*	IS*10*	1200-1350	Cut and paste	4	3	3	2	6				2				
	IS*50*	1350-1550								1	1	1	1			
	IS*H8*	1400-1800									5					
IS*5*	IS*903*	950-1150		3	5	5	19	5	1	4	4	4	4	4		
	IS*427*	800-1000		36	33	31	31	68					1			14
IS*6*	-	700-900	Conintegrate								7	1				
IS*21*	-	1750-2600									39	2	23			
IS*91*	-	1500-2000	Rolling circle								3	2				
IS*110*	IS*1111*	1200-1500		16	7	9	16	19	2	3	3	3	3		2	3
IS*200*/IS*605* [f]		1300-2000	Peel and paste	3	8	3	3	5								
	IS*200*	600-750		1	1	1	1	1	1	1				1	1	5
IS*256*	-	1200-1500	Copy and paste	104	23	28	13	47		1	1	2	2		2	1
IS*630*	-	1000-1400	Cut and paste	27	19	25	11	22								
IS*As1*	-	1200-1500		1	1		1	1	1	1	1	1	1	1	1	1
IS*L3*	-	1300-2300		30	13	12	20	20		2	3	3	3	2	3	
Tn*3*	-	>3000	Cointegrate								3	1				

Scale: 104, 68, 47, 39, 37, 31, 27, 25, 22, 19, 15, 5, 3, 2, 1

C07-087), to determine how similar they are and identify their unique features. The two pathogens cause very similar gastrointestinal septicaemia in cultured *Ictalurus* catfish in the USA, but the diseases also have unique features such as differing temperature adaptations. To extend our analysis and give a broader perspective, we included other available genomes from the NCBI genome database from the other three *Edwardsiella* species in our pan-/core-genome analyses, phylogenetic analysis and ANI analysis. Our results showed that *E. ictaluri* and *E. piscicida* are genetically distinguishable as species, but functionally conserved compared to other *Edwardsiella* species (Table 2) (Figs 1 and 2). The results also enabled correction of species designations.

Pan-/core-genome analysis of all the available sequenced *Edwardsiella* strains (as of 28/07/2017) was conducted to evaluate genomic diversity, revealing 9474 genes in the pan genome and 1421 total genes in the core genome (Fig. 1). Thus, gene acquisition and gene loss have significantly contributed to *Edwardsiella* diversification. In particular, *E. ictaluri* 93-146 has strong evidence of gene acquisition with 268 singleton proteins, which are proteins not having any orthologs in any other genome in the analysis [58]. By comparison, *E. piscicida* C07-087 has 47 singleton proteins in our analysis. Only four of the *E. ictaluri* singletons and one *E. piscicida* singleton were identified as potential virulence proteins.

ANI is the preferred digital tool for accurately estimating genome differences [59]. ANI showed that all the *Edwardsiella* species can be clearly separated from each other. Importantly, we discovered that some of the members of the genus *Edwardsiella* are not categorized properly according to ANI. For example, strains ET883, FL6-60 and EIB202 are identified as *E. tarda*, but should be classified as *E. piscicida*. These strains were classified as *E. tarda* before the relatively newer species *E. piscicida* was defined. Similarly, RSB1309 is classified as *E. piscicida*, but this designation was made prior to *E. anguillarum* being defined as a species. These strains can now be re-classified based on the current species definitions, and our results show that ANI is an effective tool to accurately classify genomes and correct misclassifications.

To be considered in the same species, the ANI should typically be 95% or more. When compared to each other, *E. anguillarum* and *E. piscicida* genomes had ANI above 95 %, making them the most closely related species in the genus. However, *E. anguillarum* is distinguishable based on ANI, and although they do not strictly meet the 95% ANI cut-off as a separate

species, it is useful to classify these strains as a separate species based on host fish species affected [59].

Phylogenetic analysis based on the complete core genomes of 29 *Edwardsiella* strains confirmed that the five *Edwardsiella* species are distinguishable from each other (Fig. 2). Similar to ANI, the core-genome comparison indicated that *E. anguillarum* and *E. piscicida* are the most closely related, and *E. ictaluri* is more closely related to these two species than *E. hoshinae* and *E. tarda*. It is interesting that phenotypically *E. tarda, E. piscicida* and *E. anguillarum* are difficult to distinguish, and *E. ictaluri* is phenotypically distinct from these three species. However, based on genome sequence comparison, *E. ictaluri* is more closely related to *E. piscicida* and *E. anguillarum* than *E. tarda* is.

Although ANI indicated that *E. piscicida* and *E. anguillarum* are the closest related *Edwardsiella* species, *E. piscicida* and *E. ictaluri* are the most functionally similar species as revealed by orthology analysis (Fig. 3). The six completely sequenced *E. ictaluri* and *E. piscicida* strains had 2495 orthologous clusters. Many of the cluster differences between the two species were in the mobilome, suggesting that acquisition of mobile elements has significantly contributed to species differentiation between *E. ictaluri* and *E. piscicida*, while functionally the two species have retained a high degree of similarity. This is apparent in the similar disease the two pathogens cause (gastrointestinal septicaemia) in *Ictalurus* catfish.

E. ictaluri had several unique features compared to the other *Edwardsiella* species. In particular, the *E. ictaluri* genomes contained urease operons that are homologous to those in *Yersinia enterocolitica* 8081 (NC_008800). In bacteria, the urease operon typically encodes a multimeric enzyme composed of three different polypeptides. Products of accessory genes are required for urease activation, and proteins encoded by *ureF*, *ureG* and *ureD* are involved in transporting nickel ions and incorporating them into the active centre of the urease apoenzyme [60]. All three are in the *E. ictaluri* genome. Urease catalyses the hydrolysis of urea to yield ammonia and carbon dioxide, thereby providing an important nitrogen source for many bacterial species, and the enzyme can contribute to the virulence of several Gram-negative bacteria by enhancing acid resistance [61]. In *E. ictaluri*, urease enables intracellular survival in professional phagocytes by mediating acid resistance, and expression of urease is induced by acidification of the vacuole [62, 63]. The three *E. ictaluri* genomes encode subunits of cytochrome o ubiquinol oxidase, which is the predominant enzyme needed for regulation of respiration associated with oxygen-rich growth conditions [64]. The *E. ictaluri* strains do not encode tetrathionate reductase operons, capsular polysaccharide synthesis enzymes and vibrioferrin-related genes that are present in the *E. piscicida* genomes, the significance of which needs further work to clarify.

Based on in-depth orthology analysis of the *E. ictaluri* 93-146 and *E. piscicida* C07-087 genomes, we focused our analysis on comparison of secretion systems and mobilome elements for these two pathogens. Secretion systems and their effector proteins contribute to the pathogenicity of members of the genus *Edwardsiella* [29, 65, 66]. However, each *Edwardsiella* species has different types and subtypes of secretion systems, revealing unique adaptations to specific environments. All of the evaluated *E. piscicida* and *E. ictaluri* strains encoded T1SS, T3SS, T5SS and T6SS. The *E. ictaluri* T3SS is necessary for virulence and intracellular replication, and effector proteins secreted into host cytoplasm have been identified [67, 68].

Encoding T4SS or some of the components of this system can give bacteria a significant advantage to adapt to environmental changes, mediate horizontal gene transfer and potentially spread antimicrobial-resistance genes [69]. None of the evaluated *E. piscicida* genomes encode a complete T4SS, but *E. piscicida* EIB2002 carries a plasmid (pEIB202) that encodes some of the T4SS elements, which contribute to transfer of multi-drug resistance genes [65]. Some of the currently available *E. piscicida* genomes are in draft form; therefore, they have not been experimentally evaluated for carrying plasmids, and it is possible that other *E. piscicida* strains carry plasmids encoding the T4SS. By contrast, four of the five currently sequenced *E. ictaluri* genomes encode the T4SS (strain RUSVM is the exception). T4SS is known for translocating DNA from other species [69]; therefore, this system may contribute to the relatively high number of mobile elements in *E. ictaluri* compared to *E. piscicida* genomes.

Three types of T6SS have been described: type I, II and III. Type I T6SS are sub-grouped into six categories, i1, i2, i3, i4a, i4b and i5, based on the conservation of core components [70–72]. All of the currently evaluated *E. ictaluri* and *E. piscicida* genomes encode T6SS type I subtype i4b, but some *E. piscicida* strains (C07-087, S11-285 and ACC35.1) also encode a second T6SS: type I subtype i2. *E. piscicida* strains ET883, FL6-60, ETW41, JF1305 and EIB202 carry only T6SS-type I subtype i4b. Interestingly, *E. anguillarum* encodes two distinct T3SS and three T6SS [3]. Thus, *Edwardsiella* species vary considerably in T4SS and T6SS. Bacterial secretion systems are potential targets for development of new anti-virulence drugs to reduce bacterial pathogenicity [73]. Therefore, further investigation of T4SS and different types of T6SS in *E. ictaluri* and *E. piscicida* is warranted.

Mobile elements contribute to horizontal gene transfer, genome organization and plasticity [50]. Three types of insertion sequences tend to be present in both *E. piscicida* and *E. ictaluri* genomes: IS*407*, IS*51* and IS*3*. IS*6*, IS*21* and IS*9I* families are present in *E. piscicida* genomes, but not *E. ictaluri*. IS*256* is present in *E. ictaluri* genomes; in *Staphylococcus aureus*, IS*256* contributes to transposon-mediated antimicrobial resistance [74]. In the panel of strains we evaluated, *E. ictaluri* genomes had a higher number of insertion sequences than *E. piscicida* genomes (Table 4). T4SS is one possible explanation, because this system is responsible for DNA and protein uptake. However, this would not explain strain RUSVM, which does not encode a complete or partial T4SS. In fact, the *E. ictaluri* RUSVM genome carries more insertion elements than any other evaluated *E. ictaluri* genome, especially in the IS*256* and IS*3* families. In *Aeromonas salmonicida*, a large number of temperature-sensitive

insertion sequences appear to contribute to genomic stability of psychrophilic strains [75]. It is interesting to speculate that the increased number of insertion sequences in *E. ictaluri*, which grows at 30 °C but not 37 °C, relative to *E. piscicida* may contribute to its genome stability.

Genomic islands contribute to acquiring virulence genes, antimicrobial-resistance genes or genes that enable adaptation to a specific environment [76]. Some *E. tarda* strains encode specific pathogenicity islands that are homologous to the genomic islands of virulent *Escherichia coli* strains [28]. In our analysis, the evaluated *E. ictaluri* genomes encode slightly more genomic islands than the *E. piscicida* genomes. Not surprisingly, a large number of the genes in these islands encode putative virulence factors; of the 512 protein-encoding genes in *E. ictaluri* 93-146 genomic islands, 212 had significant identity with known virulence genes. *E. piscicida* C07-087 is similar; 97 of 367 protein-encoding genes in genomic islands had significant similarity to known virulence proteins. This included 11 hypothetical genes and 43 phage-related genes. Thus, it is likely that genomic islands in these two species contribute to pathogenesis, and they warrant further investigation.

Phage elements are capable of mediating genetic exchange between bacteria, including transfer of virulence and antimicrobial-resistance genes. Proteins encoded in phage elements can sometimes mediate important functions such as attachment and invasion [77]. Based on our sample of evaluated genomes for *E. ictaluri* and *E. piscicida*, *E. ictaluri* strains tend to carry more phage regions. This is reflected by a higher mean number of intact prophages and total prophages per genome in the evaluated *E. ictaluri* strains compared to *E. piscicida* strains. There are five prophages integrated in the chromosome of *E. ictaluri* 93-146, whereas the *E. piscicida* C07-087 genome encodes four phage regions. It is possible that none of the *E. ictaluri* prophage regions are still active and able to excise to reproduce lytically; in one study, no temperate phage was induced by mitomycin C in 11 different *E. ictaluri* strains [78].

Integrons can play a significant role in spreading antimicrobial-resistance genes and can contribute to bacterial adaptation [53]. However, none were found in any of the *E. ictaluri* strains in the current study, and only two of the *E. piscicida* genomes (strains ETW41 and JF1305) encode some elements of integrons. Thus, integrons are not a major mechanism for chromosomally encoded antimicrobial resistance in these species.

CRISPR-Cas elements can provide protection against viral and foreign DNA [79]. These systems are not prominent in *E. ictaluri* and *E. piscicida*. Only *E. piscicida* ACC35.1 and S11-285 genomes encode a type I-E Cas system and complete CRISPR regions. Some other *E. piscicida* and *E. ictaluri* strains encode questionable CRISPR systems, but most *E. ictaluri* strains do not carry any. CRISPR-Cas systems are adaptable defence mechanisms used by many bacteria to resist predation by bacteriophage and exposure to plasmids [79, 80]. This could contribute to the *E. ictaluri* susceptibility to insertion elements and bacteriophage. There is evidence that this may be the case; in one study, *E. ictaluri* was susceptible to lysis by at least two types of bacteriophage that are specific to *E. ictaluri* [75].

In summary, comparison of the *E. ictaluri* 93-146 and *E. piscicida* C07-087 genomes has provided valuable information about the biology of these species and specific features in these catfish-adapted strains. Namely, although *E. anguillarum* and *E. piscicida* are the most closely related species in the genus based on ANI, *E. piscicida* and *E. ictaluri* are the most functionally conserved, reflecting their similar disease pathogenesis and host species. This could be the result of convergent acquisition of similar host adaptation mechanisms, but more likely the high number of mobile elements suggests that the mobilomes of these species have driven genome diversification with retention of functional pathways. Importantly, for what is believed to be the first time, we report that *E. ictaluri* genomes encode the T4SS, which could play a major role in the acquisition of mobile elements and pathogenicity of this species. *E. ictaluri* and *E. piscicida* also have strain variation in their T6SS, which could contribute to virulence and host adaptation. The relatively high number of mobile elements in *E. ictaluri* may be a result of its T4SS and lack of CRISPR-Cas systems. Overall, this comparative genomics evaluation has identified important differences between *E. ictaluri* and *E. piscicida* that warrant further study to elucidate the biology and virulence of these important fish pathogens.

Funding information
S. K. was supported by a fellowship from the Ministry of National Education of Turkey. This work was supported by the College of Veterinary Medicine at Mississippi State University, USA.

Author contributions
H.C.T., A.K. and M.L.L. designed the experiments. H.C.T., J.B., S.K. and S.N. conducted the comparative genomics analysis. The manuscript was written by H.C.T., A.K. and M.L.L. All authors read and accepted the final manuscript.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Data bibliography
All the genome sequences used in this study were downloaded from the NCBI genome database, and their GenBank accession numbers are listed in Table 1.

References

1. Ewing WH, Mcwhorter AC, Escobar MR, Lubin AH. *Edwardsiella*, a new genus of Enterobacteriaceae based on a new species, *E. tarda*. *International Bulletin of Bacteriological Nomenclature and Taxonomy* 1965;15:33–38.
2. Abayneh T, Colquhoun DJ, Sørum H. *Edwardsiella piscicida* sp. nov., a novel species pathogenic to fish. *J Appl Microbiol* 2013;114:644–654.
3. Shao S, Lai Q, Liu Q, Wu H, Xiao J *et al*. Phylogenomics characterization of a highly virulent *Edwardsiella* strain ET080813(T) encoding two distinct T3SS and three T6SS gene clusters: propose a novel species as *Edwardsiella anguillarum* sp. nov. *Syst Appl Microbiol* 2015;38:36–47.
4. Hawke JP, McWhorter AC, Steigerwalt AG, Brenner DJ. *Edwardsiella ictaluri* sp. nov., the causative agent of enteric septicemia of catfish. *Int J Syst Evol Microbiol* 1981;31:396–400.
5. Plumb JA, Sanchez DJ. Susceptibility of five species of fish to *Edwardsiella ictaluri*. *J Fish Dis* 1983;6:261–266.

6. Hawke JP, Kent M, Rogge M, Baumgartner W, Wiles J *et al*. Edwardsiellosis caused by *Edwardsiella ictaluri* in laboratory populations of zebrafish *Danio rerio*. *J Aquat Anim Health* 2013;25:171–183.
7. Soto E, Griffin M, Arauz M, Riofrio A, Martinez A *et al*. *Edwardsiella ictaluri* as the causative agent of mortality in cultured Nile tilapia. *J Aquat Anim Health* 2012;24:81–90.
8. Crumlish M, Dung TT, Turnbull JF, Ngoc NTN, Ferguson HW. Identification of *Edwardsiella ictaluri* from diseased freshwater catfish, *Pangasius hypophthalmus* (sauvage), cultured in the Mekong Delta, Vietnam. *J Fish Dis* 2002;25:733–736.
9. Sakai T, Kamaishi T, Sano M, Tensha K, Arima T *et al*. Outbreaks of *Edwardsiella ictaluri* infection in ayu *Plecoglossus altivelis* in Japanese rivers. *Fish Pathol* 2008;43:152–157.
10. Yuasa K, Kholidin EB, Panigoro N, Hatai K. First isolation of *Edwardsiella ictaluri* from cultured striped catfish *Pangasius hypophthalmus* in Indonesia. *Fish Pathol* 2003;38:181–183.
11. Liu JY, Li AH, Zhou DR, Wen ZR, Ye XP. Isolation and characterization of *Edwardsiella ictaluri* strains as pathogens from diseased yellow catfish *Pelteobagrus fulvidraco* (Richardson) cultured in China. *Aquac Res* 2010;41:1835–1844.
12. Griffin MJ, Reichley SR, Greenway TE, Quiniou SM, Ware C *et al*. Comparison of *Edwardsiella ictaluri* isolates from different hosts and geographic origins. *J Fish Dis* 2016;39:947–969.
13. Rogge ML, Dubytska L, Jung TS, Wiles J, Elkamel AA *et al*. Comparison of Vietnamese and US isolates of *Edwardsiella ictaluri*. *Dis Aquat Organ* 2013;106:17–29.
14. Griffin MJ, Ware C, Quiniou SM, Steadman JM, Gaunt PS *et al*. *Edwardsiella piscicida* identified in the southeastern USA by *gyrB* sequence, species-specific and repetitive sequence-mediated PCR. *Dis Aquat Organ* 2014;108:23–35.
15. Katharios P, Kokkari C, Dourala N, Smyrli M. First report of Edwardsiellosis in cage-cultured sharpsnout sea bream, *Diplodus puntazzo* from the Mediterranean. *BMC Vet Res* 2015;11:155.
16. Shafiei S, Viljamaa-Dirks S, Sundell K, Heinikainen S, Abayneh T *et al*. Recovery of *Edwardsiella piscicida* from farmed whitefish, *Coregonus lavaretus* (L.), in Finland. *Aquaculture* 2016;454:19–26.
17. Fogelson SB, Petty BD, Reichley SR, Ware C, Bowser PR *et al*. Histologic and molecular characterization of *Edwardsiella piscicida* infection in largemouth bass (*Micropterus salmoides*). *J Vet Diagn Invest* 2016;28:338–344.
18. Plumb JA, Quinlan EE. Survival of *Edwardsiella ictaluri* in pond water and bottom mud. *Progressive Fish Culturist* 1986;48:212–214.
19. Nusbaum KE, Morrison EE. Communications: entry of ^{35}S-labeled *Edwardsiella ictaluri* into channel catfish. *J Aquat Anim Health* 1996;8:146–149.
20. Baldwin TJ, Newton JC. Pathogenesis of enteric septicemia of channel catfish, caused by *Edwardsiella ictaluri*: bacteriologic and light and electron microscopic findings. *J Aquat Anim Health* 1993;5:189–198.
21. Morrison EE, Plumb JA. Olfactory organ of channel catfish as a site of experimental *Edwardsiella ictaluri* infection. *J Aquat Anim Health* 1994;6:101–109.
22. Shoemaker CA, Klesius PH, Plumb JA. Killing of *Edwardsiella ictaluri* by macrophages from channel catfish immune and susceptible to enteric septicemia of catfish. *Vet Immunol Immunopathol* 1997;58:181–190.
23. Miyazaki T, Plumb JA. Histopathology of *Edwardsiella ictaluri* in channel catfish, *Ictalurus punctatus* (Rafinesque)*. *J Fish Dis* 1985;8:389–392.
24. Booth NJ, Elkamel A, Thune RL. Intracellular replication of *Edwardsiella ictaluri* in channel catfish macrophages. *J Aquat Anim Health* 2006;18:101–108.
25. Chen DX, Ainsworth AJ. Assessment of metabolic activation of channel catfish peripheral blood neutrophils. *Dev Comp Immunol* 1991;15:201–208.
26. Yeh HY, Klesius PH. Changes of serum myeloperoxidase and nitric oxide in the early stage of *Edwardsiella ictaluri* infection in channel catfish, *Ictalurus punctatus* (Rafinesque). *J Fish Dis* 2013;36:441–446.
27. Yao J, Li C, Zhang J, Liu S, Feng J *et al*. Expression of nitric oxide synthase (NOS) genes in channel catfish is highly regulated and time dependent after bacterial challenges. *Dev Comp Immunol* 2014;45:74–86.
28. Nakamura Y, Takano T, Yasuike M, Sakai T, Matsuyama T *et al*. Comparative genomics reveals that a fish pathogenic bacterium *Edwardsiella tarda* has acquired the locus of enterocyte effacement (LEE) through horizontal gene transfer. *BMC Genomics* 2013;14:642–14.
29. Zheng J, Leung KY. Dissection of a type VI secretion system in *Edwardsiella tarda*. *Mol Microbiol* 2007;66:1192–1206.
30. Griffin MJ, Reichley SR, Baumgartner WA, Aarattuthodiyil S, Ware C *et al*. Emergence of *Edwardsiella piscicida* in farmed channel ♀, *Ictalurus punctatus* × Blue ♂, *Ictalurus furcatus*, hybrid catfish cultured in Mississippi. *J World Aquac Soc* 2019;50:420–432.
31. Williams ML, Gillaspy AF, Dyer DW, Thune RL, Waldbieser GC *et al*. Genome sequence of *Edwardsiella ictaluri* 93-146, a strain associated with a natural channel catfish outbreak of enteric septicemia of catfish. *J Bacteriol* 2012;194:740–741.
32. Tekedar HC, Karsi A, Williams ML, Vamenta S, Banes MM *et al*. Complete genome sequence of channel catfish gastrointestinal septicemia isolate *Edwardsiella tarda* C07-087. *Genome Announc* 2013;1:e00959-13.
33. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ *et al*. The seed and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res* 2014;42:D206–D214.
34. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP *et al*. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 2016;44:6614–6624.
35. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* 2005;102:2567–2572.
36. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P *et al*. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 2007;57:81–91.
37. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* 2009;106:19126–19131.
38. Blom J, Kreis J, Spänig S, Juhre T, Bertelli C *et al*. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res* 2016;44:W22–W28.
39. Lerat E, Daubin V, Moran NA. From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-proteobacteria. *PLoS Biol* 2003;1:e19.
40. Edgar RC. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–1797.
41. Felsenstein J. PHYLIP – phylogeny inference package (version 3.2). *Cladistics* 1989;5:164–166.
42. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. Blast ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 2011;12:402.
43. Wang Y, Coleman-Derr D, Chen G, Gu YQ. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res* 2015;43:W78–W84.
44. Abby SS, Cury J, Guglielmini J, Néron B, Touchon M *et al*. Identification of protein secretion systems in bacterial genomes. *Sci Rep* 2016;6:23080.
45. Abby SS, Rocha EPC. The non-flagellar type III secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems. *PLoS Genet* 2012;8:e1002983.
46. Guglielmini J, Néron B, Abby SS, Garcillán-Barcia MP, de la Cruz F *et al*. Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Res* 2014;42:5715–5727.

47. Li J, Yao Y, Xu HH, Hao L, Deng Z *et al*. SecReT6: a web-based resource for type VI secretion systems found in bacteria. *Environ Microbiol* 2015;17:2196–2202.
48. Bi D, Liu L, Tai C, Deng Z, Rajakumar K *et al*. SecReT4: a web-based bacterial type IV secretion system resource. *Nucleic Acids Res* 2013;41:D660–D665.
49. Lu G, Moriyama EN. Vector NTI, a balanced all-in-one sequence analysis suite. *Brief Bioinform* 2004;5:378–388.
50. Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol* 2011;12:R30.
51. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G *et al*. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res* 2017;45:W30–W35.
52. Arndt D, Grant JR, Marcu A, Sajed T, Pon A *et al*. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;44:W16–W21.
53. Cury J, Jové T, Touchon M, Néron B, Rocha EPC. Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic Acids Res* 2016;44:4539–4550.
54. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 2007;35:W52–W57.
55. Abby SS, Néron B, Ménager H, Touchon M, Rocha EPC. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS One* 2014;9:e110726.
56. Zhou CE, Smith J, Lam M, Zemla A, Dyer MD *et al*. MvirDB – a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. *Nucleic Acids Res* 2007;35:D391–D394.
57. Eikmeyer FG, Rademacher A, Hanreich A, Hennig M, Jaenicke S *et al*. Detailed analysis of metagenome datasets obtained from biogas-producing microbial communities residing in biogas reactors does not indicate the presence of putative pathogenic microorganisms. *Biotechnol Biofuels* 2013;6:49.
58. Blom J, Albaum SP, Doppmeier D, Pühler A, Vorhölter F-J *et al*. EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics* 2009;10:154.
59. Kim M, Oh H-S, Park S-C, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* 2014;64:346–351.
60. Konieczna I, Zarnowiec P, Kwinkowski M, Kolesinska B, Fraczyk J *et al*. Bacterial urease and its role in long-lasting human diseases. *Curr Protein Pept Sci* 2012;13:789–806.
61. Mobley HL, Island MD, Hausinger RP. Molecular biology of microbial ureases. *Microbiol Reviews* 1995;59:451–480.
62. Baumgartner WA, Dubytska L, Rogge ML, Mottram PJ, Thune RL. Modulation of vacuolar pH is required for replication of *Edwardsiella ictaluri* in channel catfish macrophages. *Infect Immun* 2014;82:2329–2336.
63. Booth NJ, Beekman JB, Thune RL. *Edwardsiella ictaluri* encodes an acid-activated urease that is required for intracellular replication in channel catfish (*Ictalurus punctatus*) macrophages. *Appl Environ Microbiol* 2009;75:6712–6720.
64. Cotter PA, Chepuri V, Gennis RB, Gunsalus RP. Cytochrome o (*cyoABCDE*) and d (*cydAB*) oxidase gene expression in *Escherichia coli* is regulated by oxygen, pH, and the *fnr* gene product. *J Bacteriol* 1990;172:6333–6338.
65. Hou M, Chen R, Yang D, Núñez G, Wang Z *et al*. Identification and functional characterization of EseH, a new effector of the type III secretion system of *Edwardsiella piscicida*. *Cell Microbiol* 2017;19:e12638.
66. Liu Y, Zhao L, Yang M, Yin K, Zhou X *et al*. Transcriptomic dissection of the horizontally acquired response regulator EsrB reveals its global regulatory roles in the physiological adaptation and activation of T3SS and the cognate effector repertoire in *Edwardsiella piscicida* during infection toward turbot. *Virulence* 2017;8:1355–1377.
67. Dubytska LP, Rogge ML, Thune RL. Identification and characterization of putative translocated effector proteins of the *Edwardsiella ictaluri* type III secretion system. *mSphere* 2016;1:e00039-16.
68. Thune RL, Fernandez DH, Benoit JL, Kelly-Smith M, Rogge ML *et al*. Signature-tagged mutagenesis of *Edwardsiella ictaluri* identifies virulence-related genes, including a salmonella pathogenicity island 2 class of type III secretion systems. *Appl Environ Microbiol* 2007;73:7934–7946.
69. Wallden K, Rivera-Calzada A, Waksman G. Type IV secretion systems: versatility and diversity in function. *Cell Microbiol* 2010;12:1203–1212.
70. Barret M, Egan F, Fargier E, Morrissey JP, O'Gara F. Genomic analysis of the type VI secretion systems in *Pseudomonas* spp.: novel clusters and putative effectors uncovered. *Microbiology* 2011;157:1726–1739.
71. Barret M, Egan F, O'Gara F. Distribution and diversity of bacterial secretion systems across metagenomic datasets. *Environ Microbiol Rep* 2013;5:117–126.
72. Russell AB, Wexler AG, Harding BN, Whitney JC, Bohn AJ *et al*. A type VI secretion-related pathway in Bacteroidetes mediates interbacterial antagonism. *Cell Host Microbe* 2014;16:227–236.
73. Baron C, Coombes B. Targeting bacterial secretion systems: benefits of disarmament in the microcosm. *Infect Disord Drug Targets* 2007;7:19–27.
74. Schreiber F, Szekat C, Josten M, Sahl H-G, Bierbaum G. Antibiotic-induced autoactivation of IS*256* in *Staphylococcus aureus*. *Antimicrob Agents Chemother* 2013;57:6381–6384.
75. Vincent AT, Trudel MV, Freschi L, Nagar V, Gagné-Thivierge C *et al*. Increasing genomic diversity and evidence of constrained lifestyle evolution due to insertion sequences in *Aeromonas salmonicida*. *BMC Genomics* 2016;17:44.
76. Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW *et al*. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* 2009;33:376–393.
77. Boyd EF, Carpenter MR, Chowdhury N. Mobile effector proteins on phage genomes. *Bacteriophage* 2012;2:139–148.
78. Walakira JK, Carrias AA, Hossain MJ, Jones E, Terhune JS *et al*. Identification and characterization of bacteriophages specific to the catfish pathogen, *Edwardsiella ictaluri*. *J Appl Microbiol* 2008;105:2133–2142.
79. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 2010;327:167–170.
80. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* 2014;32:347–355.
81. Reichley SR, Waldbieser GC, Soto E, Lawrence ML, Griffin MJ. Complete genome sequence of *Edwardsiella ictaluri* isolate RUSVM-1 recovered from Nile tilapia (*Oreochromis niloticus*) in the Western Hemisphere. *Genome Announc* 2017;5:e00390-17.
82. Wang R, Tekedar HC, Lawrence ML, Chouljenko VN, Kim J *et al*. Draft genome sequences of *Edwardsiella ictaluri* strains LADL11-100 and LADL11-194 isolated from zebrafish *Danio rerio*. *Genome Announc* 2015;3:e01449-15.
83. van Soest JJ, Stockhammer OW, Ordas A, Bloemberg GV, Spaink HP *et al*. Comparison of static immersion and intravenous injection systems for exposure of zebrafish embryos to the natural pathogen *Edwardsiella tarda*. *BMC Immunol* 2011;12:58.
84. Oguro K, Tamura K, Yamane J, Shimizu M, Yamamoto T *et al*. Draft genome sequences of two genetic variant strains of *Edwardsiella piscicida*, JF1305 and RSB1309, isolated from olive flounder (*Paralichythys olivaceus*) and Red Sea bream (*Pagrus major*) cultured in Japan, respectively. *Genome Announc* 2014;2:e00546-14.
85. Reichley SR, Waldbieser GC, Ucko M, Colorni A, Dubytska L *et al*. Complete genome sequence of an *Edwardsiella piscicida*-like

species isolated from diseased grouper in Israel. *Genome Announc* 2015;3:e00829-15.

86. Reichley SR, Waldbieser GC, Lawrence ML, Griffin MJ. Complete genome sequence of an *Edwardsiella piscicida*-like species, recovered from tilapia in the United States. *Genome Announc* 2015;3:e01004-15.
87. Abayneh T, Colquhoun DJ, Sørum H. Multi-locus sequence analysis (MLSA) of *Edwardsiella tarda* isolates from fish. *Vet Microbiol* 2012;158:367–375.
88. Wang Q, Yang M, Xiao J, Wu H, Wang X *et al*. Genome sequence of the versatile fish pathogen *Edwardsiella tarda* provides insights into its adaptation to broad host ranges and intracellular niches. *PLoS One* 2009;4:e7646.
89. Buján N, Toranzo AE, Magariños B. Draft genome seqGenome Sequence of *Edwardsiella piscicida* strain ACC35.1 isolated from diseased turbot (*Scophthalmus maximus*) in Europe. *Genome Announc* 2017;5:e01626-16.
90. Reichley SR, Waldbieser GC, Tekedar HC, Lawrence ML, Griffin MJ. Complete genome sequence of *Edwardsiella piscicida* isolate S11-285 recovered from channel catfish (*Ictalurus punctatus*) in Mississippi, USA. *Genome Announc* 2016;4:e01259-16.
91. Grimont PAD, Grimont F, Richard C, Sakazaki R. *Edwardsiella hoshinae*, a new species of enterobacteriaceae. *Curr Microbiol* 1980;4:347–351.
92. Reichley SR, Waldbieser GC, Lawrence ML, Griffin MJ. Complete genome sequence of *Edwardsiella hoshinae* ATCC 35051. *Genome Announc* 2017;5:e01605-16.
93. Reichley SR, Waldbieser GC, Tekedar HC, Lawrence ML, Griffin MJ. Complete genome sequence of *Edwardsiella tarda* isolate FL95-01, recovered from channel catfish. *Genome Announc* 2015;3:e00682-15.
94. Yang M, Lv Y, Xiao J, Wu H, Zheng H *et al*. *Edwardsiella* comparative phylogenomics reveal the new intra/inter-species taxonomic relationships, virulence evolution and niche adaptation mechanisms. *PLoS One* 2012;7:e36987.
95. Buján N, Lasa A, E. Toranzo A, Magariños B. Draft genome sequence of the fish strain *Edwardsiella tarda* NCIMB 2034. *Genome Announc* 2017;5:e00359-17.