Universiteit
Leiden
The Netherlands

Infectious disease studies in zebrafish : the fish pathogen Edwardsiella tarda as a model system

Soest, J.J. van

Citation
Soest, J. J. van. (2011, November 29). *Infectious disease studies in zebrafish : the fish pathogen Edwardsiella tarda as a model system*. Retrieved from https://hdl.handle.net/1887/18149

Version:	Not Applicable (or Unknown)
License:	Leiden University Non-exclusive license
Downloaded from:	https://hdl.handle.net/1887/18149

Note: To cite this publication please use the final published version (if applicable).

2

Genome comparisons of *Edwardsiella* bacteria analysed using deep sequencing technology

Joost J. van Soest, Christiaan V. Henkel, Hans J. Jansen,
Cees A.M.J.J. van den Hondel, Guido V. Bloemberg,
Annemarie H. Meijer and Herman P. Spaink

Abstract

Edwardsiella bacteria are well-known pathogens of a wide variety of animals, including humans. We have used shotgun deep sequencing technology to get a completely saturated (171 times coverage) genome sequence of *E. tarda* strain FL6-60, a common fish pathogen which is also used in zebrafish infection models. Bioinformatics analyses revealed interesting technical details of Illumina deep sequencing technology, such as a bias against regions with high CG content. We have compared our genome sequence with those of *Edwardsiella* strains already available (*E. tarda* EIB202 and *E. ictaluri* 93-146). These comparisons show that the current deep sequencing using ~50 nucleotides reads is very well suited to identify single nucleotide polymorphisms with high accuracy. Furthermore, we were able to identify all highly diversified regions that are related to transposon and viral sequences, thereby giving new insights in genome dynamics of closely related bacterial strains. The great depth of sequencing made it even possible to identify a novel circular single copy prophage that has not been found in other *Edwardsiella* strains. In addition, we show that with this technology it is possible to show with high confidence that a plasmid present in *E. tarda* strain EIB202 is not present in strain FL6-60. As an example of the usefulness of these datasets we compared single nucleotide polymorphisms in several gene clusters that are associated with virulence, also using other published incomplete *Edwardsiella* sequences. The results showed that several virulence genes, including those coding for type III and type VI secretion systems, have a remarkable high number of polymorphisms. The identification of these variations will be important for further comparisons in pathogenicity and virulence of different *Edwardsiella* strains.

Introduction

Edwardsiella tarda is a gram negative pathogen capable of infecting a wide range of host species, such as fish, amphibians, reptiles, birds and mammals, including humans [1, 2]. Edwardsiellosis, a generalized septicemia caused by *E. tarda*, is an often occurring disease in many fish species in aquaculture, which leads to extensive losses [3]. In humans, *E. tarda* is also a cause of

gastrointestinal and extra-intestinal infections [1] and is able to invade non-phagocytic cells in culture [4, 5].

E. tarda strain FL6-60 was shown to be able to infect zebrafish embryos by static immersion [6]. Adult zebrafish were susceptible to intraperitoneal injection and could also be infected by static immersion, but only in combination with skin wounding. With the zebrafish being increasingly used as a model system for infectious diseases [7], this makes this strain a valuable test organism for disease modelling. The zebrafish *E. tarda* infection model is now also used for detailed innate immune transcriptome analysis (this thesis, chapters 3 and 4). Recently the complete genome sequence of *Edwardsiella tarda* strain EIB202 has been published [8]. EIB202 was shown to be pathogenic in swordfish, turbot and adult zebrafish, but this strain has not been tested yet in zebrafish embryo infection models. Furthermore, the genome of its close relative *E. ictaluri* strain 93-146 has been completely sequenced and submitted to the Genbank database (accession number CP001600). In addition, a collection of unassembled short fragment reads of a shotgun sequencing attempt of *E. tarda* strain ATCC 23685 is available (SRA accession numbers SRX001436 and SRX001437).

The development of new sequencing techniques has rapidly reduced the time and cost of whole-genome sequencing making it attractive to compare many bacterial genomes [9]. The next-generation sequence platforms, provided by Illumina, ABI and 454 Life Sciences (Roche), generate large amounts of nucleotide sequence reads [10]. These techniques were primarily developed for resequencing of closely related individuals, but are with the increase of the read sequence lengths and the availability of assembly software specialized for short reads gradually becoming used for *de novo* genome sequencing [11-14]. We sequenced *E. tarda* strain FL6-60 with the Illumina Genome Analyzer II, which provided us with reads of a length of 51 bp that were assembled into large contigs and compared to the EIB202 sequence. We have used the FL6-60 genome sequence for comparison of virulence genes with the known sequences of other *E. tarda* strains to see if there are genetic differences between strains that can be a basis for differences in pathogenic behaviour. Several virulence factors, common in Gram-negative pathogens, have been identified as being important in *E. tarda* virulence. The three major systems involved are the protein secretion systems Type III Secretion System (T3SS)

[15] and Type VI Secretion System (T6SS) [16] and the regulatory quorum sensing system [17, 18].

The virulence determinants of *E. tarda* have been identified by functional genomics approaches such as transposon mutagenesis [19] and proteomics [20]. Of several *E. tarda* strains some of the genes found to be associated with virulence were sequenced, for instance the complete sequences are available for the T3SS gene cluster [15], the T6SS gene cluster [16], the quorum sensing genes LuxIR and LuxS, and some genes involved in toxicity [19] and resistance to phagocyte-mediated killing [19, 21, 22].

Our sequence comparisons show that the used current deep sequencing technology is highly suited to identify single nucleotide polymorphisms with high accuracy. For two highly related bacterial strains, we have mapped genome dynamic regions and identified all single nucleotide polymorphisms in the above mentioned loci and the presence or absence of plasmids or circular prophage-related sequences.

Results and discussion

Genome assembly strategy

The genome of *E. tarda* FL6-60 was assembled using a combination of automated *de novo* assembly, reference assembly against the recently published genome of *E. tarda* EIB202 and manual inspection of highly variable regions (see table 1). Using *de novo* assembly alone, 98.2% of the sequencing reads could be assigned to contigs of at least 200 bp. In a reference assembly using the *E. tarda* EIB202 genome, 92.4% of reads could be mapped. On average, every position on this genome is covered by 171 reads; however in total 145443 bp (3.9% of the genome) could not covered by any read. A reference assembly using plasmid pEIB202 of strain EIB202 did not yield a relevant number of matching reads, thereby demonstrating the absence of this plasmid in strain FL6-60.

In the reference assembly, many regions of the EIB202 genome are uniquely matched by sequencing reads with at least twice the average coverage (figure 1), indicating possible genomic duplications. However, when comparing

Table 1. Genome statistics of *Edwardsiella tarda* strain FL6-60.

Assembly	Reference	Accession	Length	%CG	ORFs*	Assembled reads	Average coverage
De novo	-	-	3801681 bp in 623 contigs (N50 = 12020 bp)	ND	ND	145153318 (98.2%)	180.16x
Reference	*E. ictaluri* 93-146 genome	NC_012779	3812315 bp	57.4%	3935	5234702 (35.4%)	63.47x
Reference	*E. tarda* EIB202 genome	NC_013508	3760463 bp	59.7%	3664	13661894 (92.4%)	171.01x
Reference	*E. tarda* pEIB202 plasmid	NC_013509	43703 bp	57.3%	53	2 (0.0%)	0.00x
Reference	*E. tarda* FL6-60 genome		3684607 bp	59.8%	3448	14250618 (96.4%)	182.38x
Reference	*E. tarda* FL6-60 phage-like element		44194 bp	51.4%	63	268143 (1.8%)	293.07x

* The number of ORFs in the reference sequence as found in Genbank or determined by Glimmer

local coverage to GC-content, a clear correlation was found (figure 2), suggesting a technical artefact that we have further analysed. A bias against AT-rich regions has been previously reported for Illumina sequencing [23]. In apparent contrast, we find a strong correlation between GC-content and sequencing depth: the lower the GC-content of the reference, the higher the number of matching reads. However, the *E. tarda* EIB202 genome contains very few AT-rich regions even near the resolution of single sequencing reads (100 bp windows), so any bias against AT-rich sequences would be difficult to observe. In fact, the observed apparent linear relationship between GC% and coverage (figure 1) makes it possible to predict sequencing depth for the whole reference genome, allowing several outliers to be quickly identified.

A complete draft genome of strain FL6-60 was then constructed by manually integrating *de novo* assembled contigs into the gaps of the reference assembly using matching flanking sequences (figure 2). One large contig could

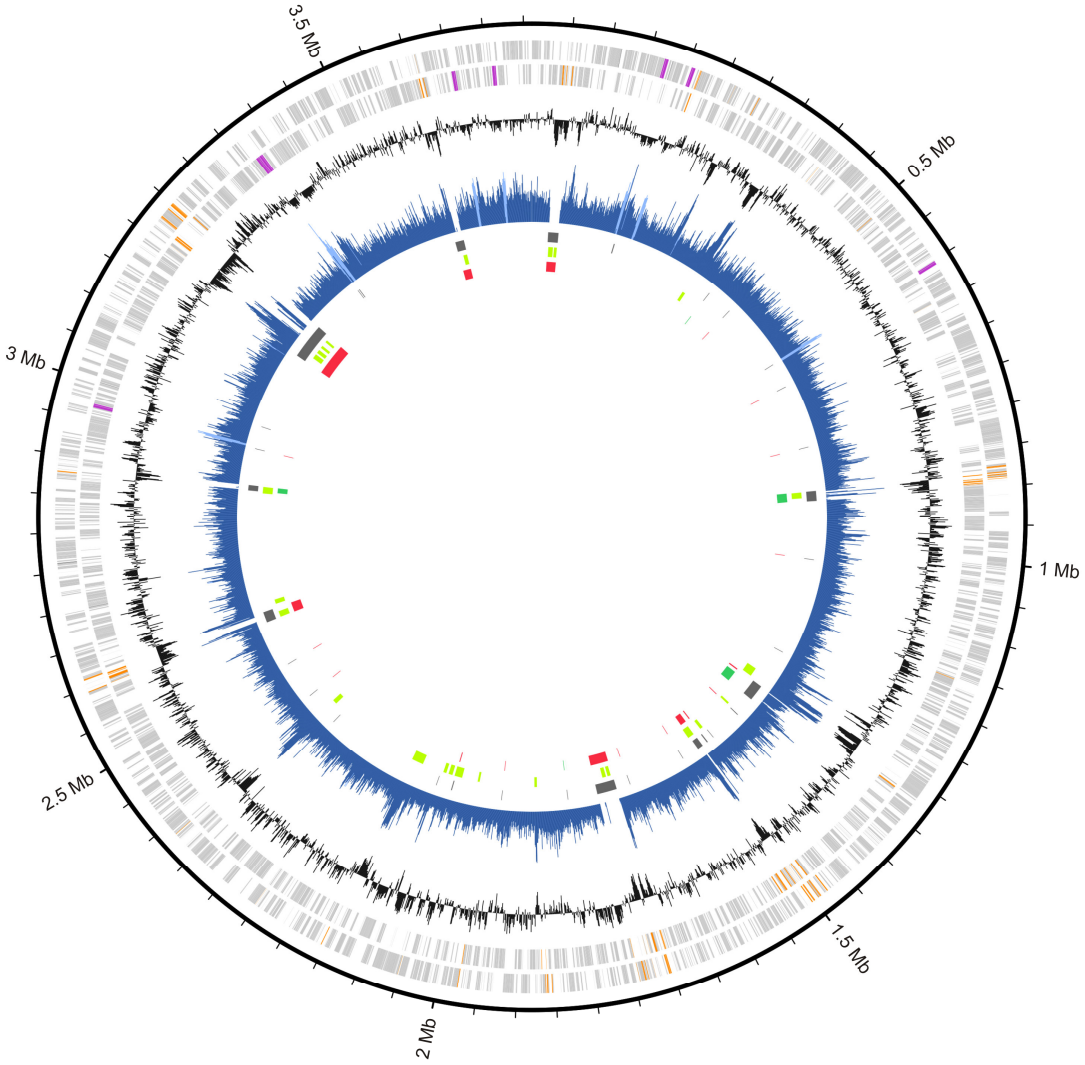

Figure 1. Overview of the genome of *Edwardsiella tarda* strain EIB202 excluding one circular phage-like element. The circles from outside to inside represent: Annotated genes plus strand (grey, protein coding; black, tRNA; purple, rRNA; orange, phage-related/integrase/transposase). Annotated genes minus strand (colour coding same as plus strand). CG content above or below the median (calculated for 1000 bp windows). Sequencing depth (coverage) for the reference assembly of strain FL6-60 (1000 bp windows), dark blue are unique matches, light blue are non-unique matches of sequencing reads. Regions masked in the second iteration of the reference assembly of strain FL6-60 (dark grey). Genomic islands as determined by IslandViewer (lime). Net insertions (green) and deletions (red) in strain FL6-60 with regard to strain EIB202, genes affected can be found in supplementary table S1.

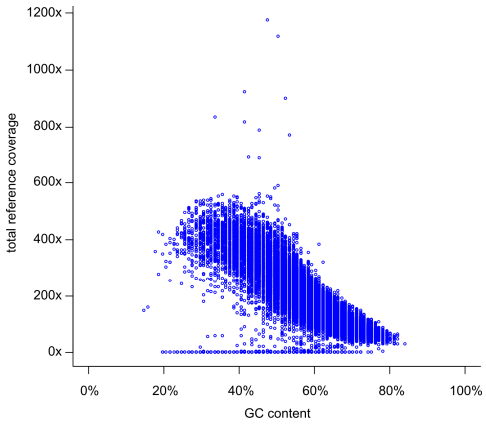

Figure 2. GC bias of Illumina sequencing. GC content and average coverage were calculated for 100 bp windows of the EIB202 genome. For 40-80% GC, the relationship with read coverage can be described by a straight line, allowing easy identification of regions with aberrant coverage. As there are few 100 bp windows with less than 25% GC, this relationship does not contradict reports of a bias against AT-rich regions.

Figure 3. Map of a 44194 bp circular phage-like element of *Edwardsiella tarda* strain FL6-60. The circles from outside to inside represent: Scale; Annotated genes (orange is phage related), numbers with the genes indicate the ORF numbers, which correspond to those in table 2, a more detailed table can be found in supplementary table S2; GC content above or below median (1000 bp windows); Sequencing depth (coverage, 1000bp windows).

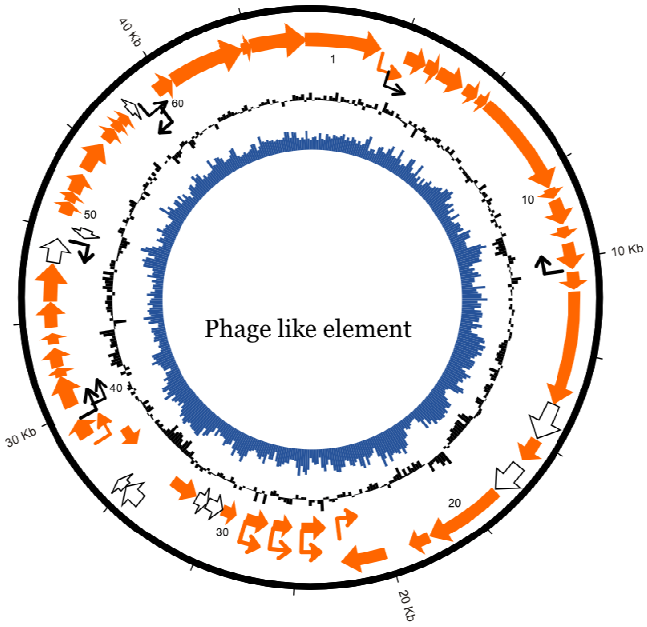

Table 2. Reading frame annotation of a circular phage-like element of *Edwardsiella tarda* FL6-60

Predicted functions based on sequence similarity	Number of open reading frame*
Phage tail components	4-11, 13, 15, 20, 21
Phage membrane proteins	12, 18, 50
Phage methyltransferases	30, 43
Other phage related proteins	1, 16, 22, 23, 26, 33, 36, 37, 41, 44, 51-55,
(i.e. nucleases, integrases, repressors)	60, 61, 63
Phage hypothetical proteins	2, 24, 25, 28, 29, 33, 38, 42, 45, 46, 56, 62
other proteins	35, 58
hypothetical proteins	3, 17, 19, 27, 31, 32, 34, 39, 40, 47, 49, 57,
no blast hit	14, 48, 59

* as indicated in figure 3

not be assigned to the genome, and indeed appears to be a separate circular entity (figure 3). In a new reference assembly, 98.2% of reads (99.6% using less stringent settings) matched either the new draft genome or the circular element. Since slightly more reads match the drafts than can be assembled *de novo* (table 1), and the remaining 1.8% could not be further assembled *de novo* into significant contigs because of sequence repeats, we accepted the draft as an essentially correct representation of the *E. tarda* FL6-60 genome.

Comparison of *E. tarda* FL6-60, *E. tarda* EIB202 and *E. ictaluri* 93-146

The genome of *E. tarda* strain FL6-60 is highly similar to that of strain EIB202 as previously described by Wang et al [8]. Genomic differences due to insertions and deletions can be found in the supplementary data but these are limited to areas containing many phage-related sequences (table S1). Nearly 3300 reading frames annotated for the genome of EIB202 by Wang et al [8] could also be identified in the genome of strain FL6-60. Automatic sequence annotation did not reveal any particular gene differences with the published sequence of Wang et al [8] (data not shown). We have looked in detail to several gene clusters not discussed in detail by Wang et al [8] such as the genes involved in exopolysaccharide production and found these genes to be highly conserved as well (99.9-100 % nucleotide sequence identity) (data not shown). Differences in the genome represent recent evolutionary changes related to phage integrations and shuffling of repetitive sequences. The largest differences between the strains appear to be concentrated around predicted genomic islands, which harbour phage-related genes and transposons (figure 1).

A major difference is the absence of a plasmid that is present in strain EIB202. In contrast strain FL6-60 contains an additional circular element (figure 3). This element has a similar length as representatives of the P22 podoviridae and contains many predicted transcripts encoding proteins that are homologous to phage proteins (table 2). This element possibly represents a novel viral element. Because of the correlation between sequencing depth and GC-content, the higher coverage and lower GC-content in comparison with the genome, indicates that this element is present in a copy number of 1 per cell,

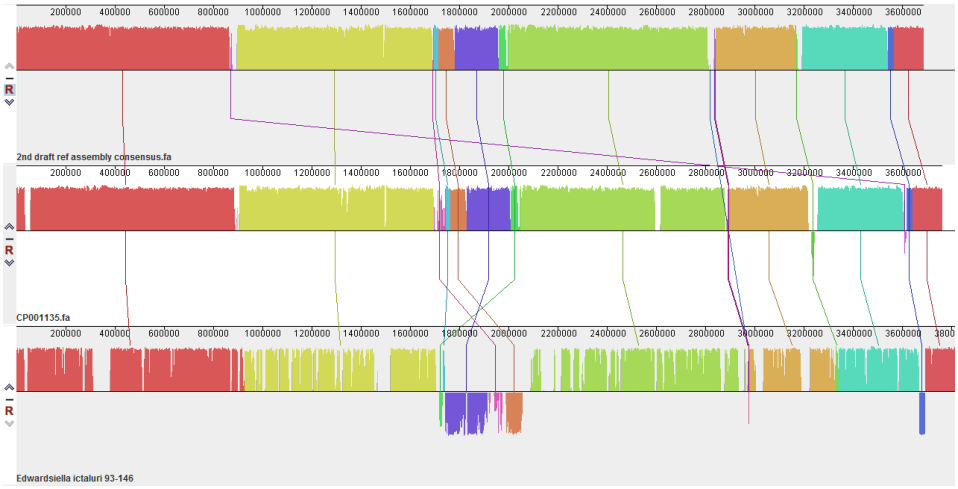

Figure 4. Comparison of the genomes of *Edwardsiella tarda* strains FL6-60 and EIB202 with *E. ictaluri* strain 93-146. The genomes of *Edwardsiella tarda* strains FL6-60 and EIB202 were compared with *E. ictaluri* strain 93-146 with Mauve. The graphs indicate the conservation. Conserved areas are connected with lines between the different genomes, identifying inversions and displacements.

indicating a strict co-replication with the chromosome. We have also compared the genome of strain FL6-60 with that of *E. ictaluri* strain 93-146 (figure 4). The results show there is remarkable overall structural similarity with 35.4 % of identity (table 1) and only two regions with a different orientation as confirmed by manual inspection of the variable regions (figure 4).

Comparison of type III and type VI secretion systems between different *Edwardsiella* strains

It was found that the pathogenicity of different natural isolates of *E. tarda* was corresponding with the presence of the T6SS since strains lacking the T6SS were not pathogenic [16]. Likewise, *E. tarda* strains with mutations that partly or completely shut down the T3SS were less virulent then their wild types [15].

Since type III and VI secretion systems are important virulence factors, knockdown of a single gene can already cause a decrease in pathogenicity. Small changes in these secretion systems might lead to loss of virulence or changing the specialization for a specific host. We have therefore compared these secretion systems in detail for five type III and four type VI secretion systems of *Edwardsiella*. The results show that the secretion systems of strains

FL6-60 and EIB202 are more similar than those of the other *E. tarda* strains (figure 5 and 6). Even though in *E. ictaluri* the T3SS and T6SS clusters are present and in the same order of genes as in *E. tarda*, there are large differences in the encoded protein sequences between the two *Edwardsiella* species. This might be explained by the fact that *E. ictaluri* is primarily pathogenic to catfish and therefore might have specialized for this host, in contrast to *E. tarda* that appears to be less host-specific. It would therefore be

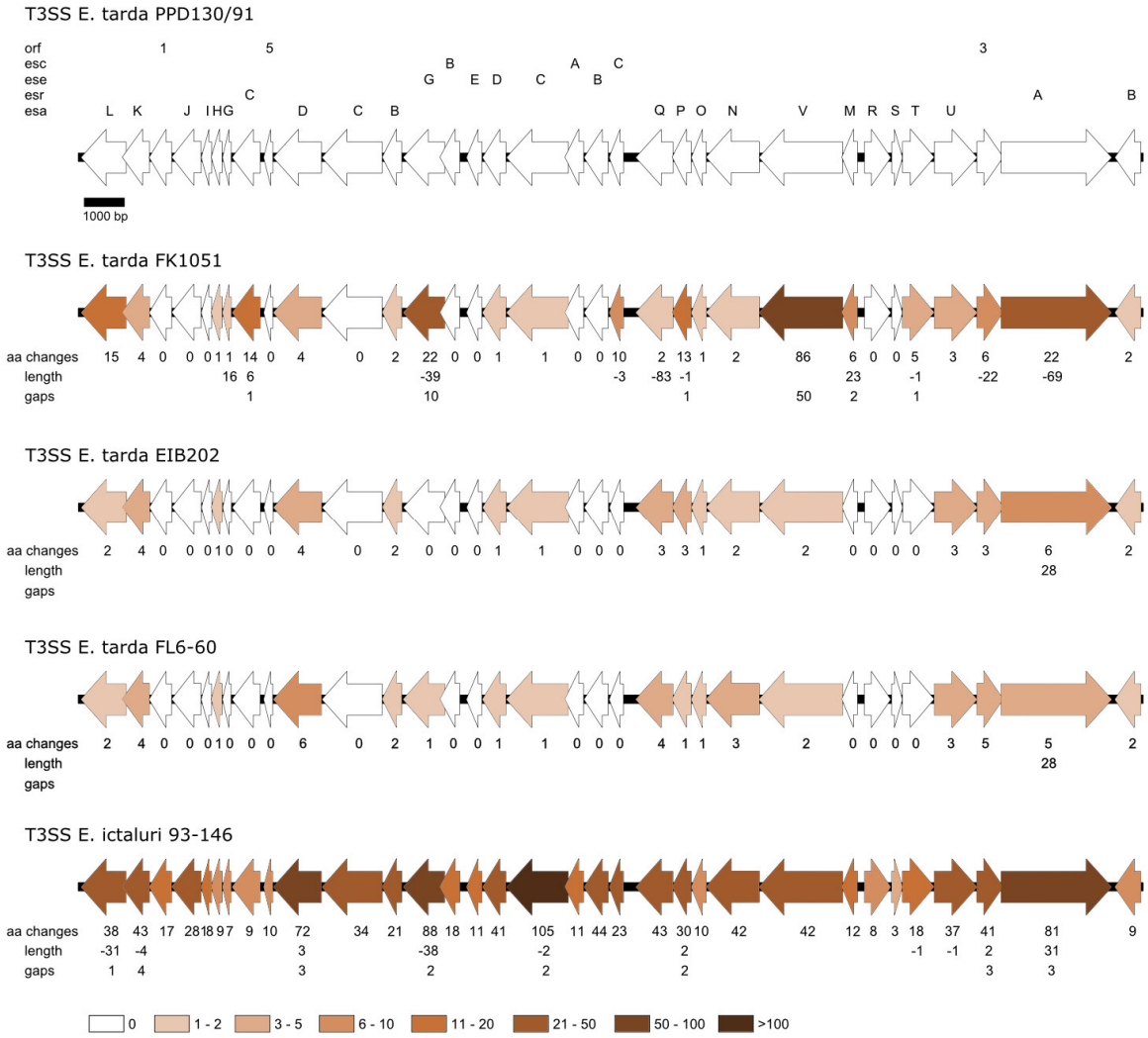

Figure 5. Comparisons of the type III secretion systems of five different *Edwardsiella* strains. The T3SS of *E. tarda* PPD130/91 [15] was used as the reference for comparison. The differences in amino acids in comparison with this strain are notated underneath the clusters. A darker colour represents a larger difference. To determine the difference in length, the nucleotide sequences were compared first, to exclude differences in length due to different ORF-predictions. The order of the genes is the same in all strains compared, with the largest differences found in *E. ictaluri*.

interesting to compare these strains in the same test systems, along with several mutants in the genes that are differing between these strains. The gene that is most different in the type III secretion systems of strains FL6-60 and EIB202 is *esrA* which is a regulatory gene of a two component system together with *esrB*. Other genes that differ considerably are part of the secretion apparatus, such as *esaD* (figure 5). There are many differences with the type III secretion system of *E. tarda* strain FK1051, with notable large differences in the genes *eseG* and *esaV* that are secreted proteins or involved in the secretion machinery, respectively. It remains a possibility that many of the apparent polymorphisms detected are due to sequencing errors in the previously submitted sequences, however the degree of polymorphism is too much to

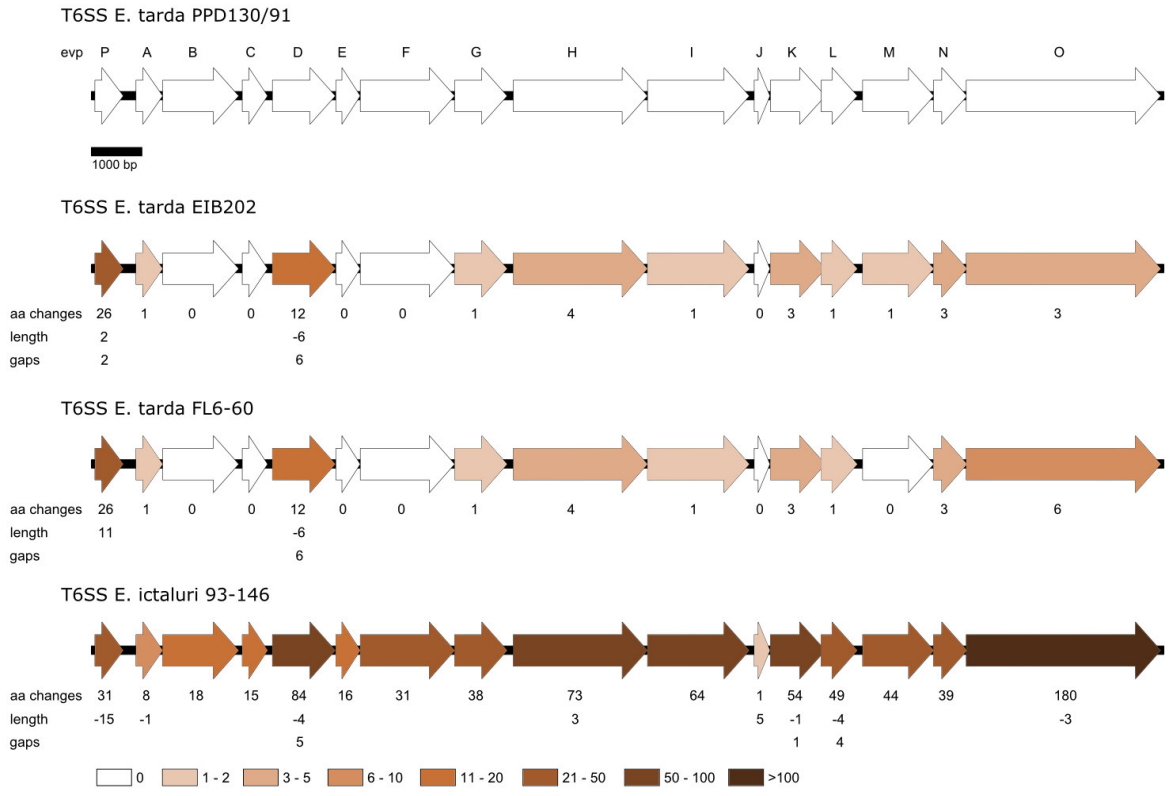

Figure 6. Comparisons of the type VI secretion systems of four different *Edwardsiella* strains. The T6SS of *E. tarda* PPD130/91 [16] was used as the reference for comparison. The differences in amino acids in comparison with this strain are notated underneath the clusters. A darker colour represents a larger difference. To determine the difference in length, the nucleotide sequences were compared first, to exclude differences in length due to different ORF-predictions. The order of the genes is the same in all strains compared, with the largest differences found in *E. ictaluri*.

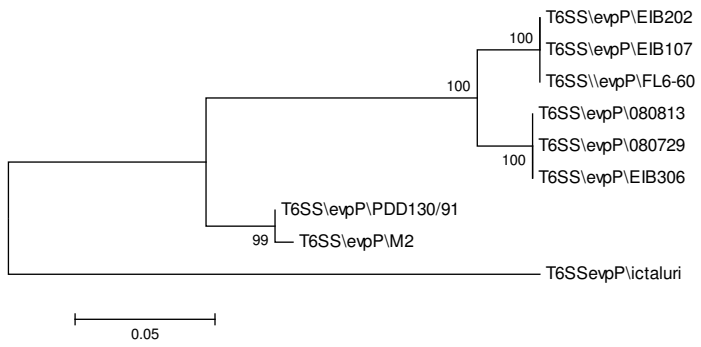

Figure 7. Evolutionary relationships of 9 EvpP genes. The evolutionary history was inferred using the Neighbour-Joining method [25]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches [26]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

be accounted for by sequencing errors. The most obvious difference with the genes of *E. ictaluri* is in *eseC* for which 31 of the 105 predicted amino acid differences are localized in an area that has been shown to be important for secretion and accumulation of EseC in *E. tarda* [24]. Furthermore, the large differences between *eseB*, *eseC* and *eseD* are perhaps relevant since these genes have been shown to be important for virulence of *E. tarda* [15]. For the type VI secretion systems the difference between the *evpP* genes that are unique to *Edwardsiella* and are involved in pathogenicity [16] is shown in more detail for 9 *Edwardsiella* strains. The comparison (figure 7), shows that *evpP* of *E. ictaluri* is most different from the *E. tarda* strains and therefore just following the expected species divergence and giving no indications for recent horizontal gene transfer events between these species.

Comparison of non-essential genes showed that there is no clear correlation of the conservation in protein sequence with the essentiality of the gene for secretion as for instance *evpJ* is strongly conserved in contrast to *evpD*, whereas both have been shown not to be essential for secretion [16].

Concluding remarks

Our comparisons show that the current deep sequencing using 50 nucleotides reads is highly suited to identify an entire genome with high accuracy. Bioinformatics analyses also showed interesting technical details of deep sequencing technologies such as depletion of GC rich regions by the Illumina technology.

As a result we were able to compare the genome of *E. tarda* FL6-60 with previously published sequences of other strains and could identify all highly diversified regions with many repeats that are related to transposon and viral sequences. These data provide new insights in genomic variations of highly related bacterial strains. We have identified one novel circular phage-like element and could show with high confidence that a plasmid present in *E. tarda* strain EIB202 was not present in strain FL6-60. As an example of the usefulness of gene comparisons we investigated single nucleotide polymorphisms in several gene clusters that are associated with virulence also using other published incomplete *Edwardsiella* sequences. The results showed that several of the type III and type VI secretion genes, most notably *eseG*, *esaV*, *esrA* and *evpP* have a remarkable high number of polymorphisms not only between *E. tarda* and *E. ictaluri* species but also between different *E. tarda* strains. Other known virulence factors such as enzymes involved in resistance to phagocyte-mediated killing, like *katB* and *ankB* [19, 21, 22] showed only 11 SNPs difference between strains FL6-60 and EIB202. The quorum sensing system was found to be almost completely identical (2 SNPs) to those already published [17, 18]. The identification of these genomic variations in essential virulence genes provides a valuable basis for studying the pathogenic behaviour of different *Edwardsiella* strains.

Materials & methods

Sequencing

Bacterial strain *E. tarda* FL6-60 was grown on TSB medium at 28 °C. Bacterial DNA was isolated using the Qiagen DNeasy Blood and Tissue kit (Qiagen, San Diego, CA) according to the manufacturer's instructions. To sequence the DNA, a single read library was made using the Illumina Genomic DNA Sample Preparation Kit (Illumina, Hilden, Germany) according to the manufacturer's instructions. This library was sequenced, with a read length of 51 nt, on an Illumina GAII (Illumina, Hilden, Germany). Two lanes were sequenced, one lane with 3 pmol , and one lane with 4 pmol of library. A total of 16606657 sequencing reads was obtained.

Bioinformatics

Sequence reads were imported, trimmed to eliminate low quality sequences, and assembled using CLCbio Genomics Workbench 3.6.5 (www.clcbio.com). Putative genes were found using Glimmer3 (www.ncbi.nlm.nih.gov/genomes/MICROBES/glimmer_3.cgi) [27]. Genomic islands were predicted with IslandViewer (www.pathogenomics.sfu.ca/islandviewer) [28], which combines several methods based on sequence composition and comparative genomics.

Comparison of our assembled genome with the available genomes of *E. tarda* (accession number CP001135) and *E. ictaluri* (accession number CP001600) was done with Mauve 2.3.1 [29]. Gene analysis was done with Vector NTI and BLAST on the NCBI server.

Circos [30] was used for the construction of the maps of the genome and the circular phage-like element. Mega4 was used for the construction of the phylogenetic tree [31]. The evolutionary distances were computed using the Poisson correction method [32] and are given in the units of the number of amino acid substitutions per site.

Sequences used in comparisons were obtained from the NCBI-database with the following accession numbers excluding the ones already mentioned. T3SS: AY643478, AY850613. T6SS: AY424360. EvpP: FJ595672, FJ595674, FJ595675, FJ595676, FJ595677

Accession numbers

The genome sequence has been submitted under accession number CP002154 and the sequence of the phage-like element has been submitted under accession number CP002155.

Acknowledgements

We thank Dr. Philip Klesius (USDA, Auburn, AL) for providing us with *E. tarda* strain FL6-60. The research group of C.A.M.J.J.H. is part of the Kluyver Centre for Genomics of Industrial Fermentation, which is supported by the Netherlands Genomics Initiative.

Supplementary data

Supplementary tables can be found online at: http://tinyurl.com/ch2suptables

References

1 Janda, J. M. and S.L. Abbott. 1993. Infections associated with the genus Edwardsiella: the role of *Edwardsiella tarda* in human disease. Clin Infect Dis 17:742-748.

2 Kourany, M., M.A.Vasquez and R. Saenz. 1977. Edwardsiellosis in man and animals in Panama: clinical and epidemiological characteristics. Am J Trop Med Hyg 26:1183-1190.

3 Thune, R. L., L. Stanley, and R.K. Cooper. 1993. Pathogenesis of Gram-negative bacterial infections in warmwater fish. Fish Annu Rev Fish Dis 3:37-68

4 Ling, S. H., X.H. Wang, T.M. Lim, and K.Y. Leung. 2001. Green fluorescent protein-tagged *Edwardsiella tarda* reveals portal of entry in fish. FEMS Microbiol Lett 194:239-243.

5 Ling, S. H. M., X.H. Wang, L. Xie, T.M. Lim, and K.Y. Leung. 2000. Use of green fluorescent protein (GFP) to study the invasion pathways of *Edwardsiella tarda* in *in vivo* and *in vitro* fish models. Microbiology 146:7-19.

6 Pressley, M. E., P.E. Phelan III, P.E. Witten, M.T. Mellon, and C.H. Kim. 2005. Pathogenesis and inflammatory response to *Edwardsiella* tarda infection in the zebrafish. Dev Comp Immunol 29:501-513.

7 Stockhammer, O. W., A. Zakrzewska, Z. Hegedus, H.P. Spaink, and A.H. Meijer. 2009. Transcriptome profiling and functional analyses of the zebrafish embryonic innate immune response to salmonella infection. J Immunol 182:5641-5653.

8 Wang, Q., M. Yang, J. Xiao, H. Wu, X. Wang, Y. Lv, L. Xu, H. Zeng, S. Wang, G. Zhao, Q. Liu, Y. Zhang. 2009b. Genome sequence of the versatile fish pathogen *Edwardsiella tarda* provides insights into its adaptation to broad host ranges and intracellular niches. PLoS One 4:e7646.

9 Wu, D., P. Hugenholtz, K. Mavromatis, R. Pukall, E. Dalin, N.N. Ivanova, V. Kunin, L. Goodwin, M. Wu, B.J. Tindall, S.D. Hooper, A. Pati, A. Lykidis, S. Spring, I.J. Anderson, P. D'haeseleer, A. Zemla, M. Singer, A. Lapidus, M. Nolan, A. Copeland, C. Han, F. Chen, J.F. Cheng, S. Lucas, C. Kerfeld, E. Lang, S. Gronow, P. Chain, D. Bruce, E.M. Rubin, N.C. Kyrpides, H.P. Klenk, and J.A. Eisen. 2009. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature 462:1056-1060.

10 Margulies, M., M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt , G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley, and J.M. Rothberg.. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380.

11 Hogg, J., F. Hu, B. Janto, R. Boissy, J. Hayes, R. Keefe, J.C. Post, and G. Ehrlich. 2007. Characterization and modelling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. Genome Biol 8:R103.

12 McCutcheon, J. P. and N.A. Moran. 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proc Natl Acad Sci USA 104:19392-19397.

13 Reinhardt, J. A., D.A. Baltrus, M.T. Nishimura, W.R. Jeck, C.D. Jones, and J.L. Dangl. 2009. *De novo* assembly using low-coverage short read sequence data from the rice pathogen *Pseudomonas syringae* pv. *oryzae*. Genome Res 19:294-305.

14 Worley, K. C. and R.A. Gibbs. 2010. Genetics: Decoding a national treasure. Nature 463:303-304.

15 Tan, Y. P., J. Zheng, S.L. Tung, I. Rosenshine, and K.Y. Leung. 2005. Role of type III secretion in *Edwardsiella tarda* virulence. Microbiology 151:2301-2313.

16 Zheng, J. and K.Y. Leung. 2007. Dissection of a type VI secretion system in *Edwardsiella tarda*. Mol Microbiol 66:1192-1206.

17 Zhang, M., K. Sun, and L. Sun. 2008. Regulation of autoinducer 2 production and luxS expression in a pathogenic *Edwardsiella tarda* strain. Microbiology 154:2060-2069.

18 Morohoshi, T., T. Inaba, N. Kato, K. Kanai, and T. Ikeda. 2004. Identification of quorum-sensing signal molecules and the LuxRI homologs in fish pathogen *Edwardsiella tarda*. J Biosci Bioeng 98:274-281.

19 Srinivasa Rao, P. S., T.M. Lim, and K.Y. Leung. 2003. Functional genomics approach to the identification of virulence genes involved in *Edwardsiella tarda* pathogenesis. Infect Immun 71:1343-1351.

20 Tan, Y. P., Q. Lin, X.H. Wang, S. Joshi, C.L. Hew, and K.Y. Leung. 2002. Comparative proteomic analysis of extracellular proteins of edwardsiella tarda. Infect Immun 70:6475-6480.

21 Srinivasa Rao, P. S., Y. Yamada, and K.Y. Leung. 2003. A major catalase (KatB) that is required for resistance to H2O2 and phagocyte-mediated killing in *Edwardsiella tarda*. Microbiology 149:2635-2644.

22 Han, H. J., D.H. Kim, D.C. Lee, S.M. Kim, and S.I. Park. 2006. Pathogenicity of *Edwardsiella tarda* to olive flounder, *Paralichthys olivaceus* (Temminck & Schlegel). J Fish Dis 29:601-609.

23 Hillier, L.W., G.T. Marth, A.R. Quinlan, D. Dooling, G. Fewell, D. Barnett, P. Fox, J.I. Glasscock, M. Hickenbotham, W. Huang, V.J. Magrini, R.J. Richt , S.N. Sander, D.A. Stewart, M. Stromberg, E.F. Tsung, T. Wylie, T. Schedl, R.K. Wilson, E.R. Mardis. 2008. Whole-genome sequencing and variant discovery in *C. elegans*. Nat Meth 5:183-188.

24 Wang, B., Z.L. Mo, Y.X. Mao, Y.X. Zou, P. Xiao, J. Li, J.Y. Yang, X.H. Ye , K.Y. Leung, P.J. Zhang. 2009. Investigation of EscA as a chaperone for the *Edwardsiella tarda* type III secretion system putative translocon component EseC. Microbiology 155: 1260-1271.

25 Saitou, N. and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425.

26 Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.

27 Delcher, A., D. Harmon, S. Kasif, O. White, and S. Salzberg. 1999. Improved microbial gene identification with GLIMMER. Nucl Acids Res 27:4636-4641.

28 Langille, M. G. I. and F.S.L. Brinkman. 2009. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. Bioinformatics 25:664-665.

29 Darling, A. C. E., B. Mau, F.R. Blattner, and N.T. Perna. 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394-1403.

30 Krzywinski, M., J. Schein, Ä.n. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, and M.A. Marra. 2009. Circos: An information aesthetic for comparative genomics. Genome Res 19:1639-1645.

31 Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol Biol Evol 24:1596-1599.

32 Zuckerkandl, E., L. Pauling, V. Bryson, and H.J. Vogel. 1965. Evolutionary divergence and convergence in proteins. In Evolving Genes and Proteins, pp. 97-166.