Taxonomic Description template

Escherichia ruysiae sp. nov., isolated from an international traveller

Boas C.L. van der Putten[1,2#], S. Matamoros[1], D.R. Mende[3], COMBAT consortium[†], C. Schultsz[1,2]

[1]Department of Medical Microbiology, Amsterdam UMC, University of Amsterdam, the Netherlands

[2]Department of Global Health, Amsterdam Institute for Global Health and Development, Amsterdam UMC, University of Amsterdam, the Netherlands

[3]Department of Applied Evolutionary Biology, Amsterdam UMC, University of Amsterdam, the Netherlands

[#]Corresponding author, email address: boas.vanderputten@amsterdamumc.nl

[†]Members listed in supplementary file 1

Keywords: Escherichia, bacterial taxonomy, Enterobacteriaceae

ABSTRACT

The *Escherichia* genus comprises four species and at least five lineages currently not assigned to any species, termed '*Escherichia* cryptic clades'. We isolated an *Escherichia* strain from an international traveller and resolved the complete DNA sequence of the chromosome and an IncI multi-drug resistance plasmid using Illumina and Nanopore whole-genome sequencing (WGS). Strain OPT1704^T can be differentiated from existing *Escherichia* spp. using biochemical (VITEK2) and genomic tests (average nucleotide identity [ANI] and digital DNA:DNA hybridisation [dDDH]). Phylogenetic analysis based on alignment of 16S rRNA sequences and 682 concatenated core genes showed similar results. Our analysis further revealed that strain OPT1704^T falls within *Escherichia* cryptic clade IV, and is closely related to cryptic clade III. Combining our analyses with publicly available WGS data of

cryptic clades III and IV from Enterobase confirmed the close relationship between clades III and IV (>96% interclade ANI), warranting assignment of both clades to the same novel species. We propose *E. ruysiae* sp. nov. as a novel species, encompassing *Escherichia* cryptic clades III and IV (type strain OPT1704^T = NCCB 100732^T = NCTC 14359^T).

Author notes

The Genbank accession number for the 16S rRNA gene sequence of strain OPT1704^T is LR745848.

The Genbank accession number for the complete genome sequence of strain OPT1704^T is CABVLQ000000000.

Introduction

Within the *Escherichia* genus, four species are recognized; *E. coli* (1), *E. fergusonii* (2), *E. albertii* (3) and most recently, *E. marmotae* (4). Several species were assigned to the *Escherichia* genus previously, but have now been moved to other genera, such as *E. vulneris* (now *Pseudescherichia vulneris* (5)), *E. blattae* (now *Shimwellia blattae* (6)), *E. adecarboxylata* (now *Leclercia adecarboxylata* (7)) and *E. hermannii* (now *Atlantibacter hermannii* (8)). All four *Escherichia* species have been associated with the potential to cause animal and/or human disease (9–12). Several *Escherichia* strains cannot be assigned to any of the four existing species. Based on analysis of genetic data, these strains cluster into several groups, which were termed '*Escherichia* cryptic clades' (13,14). Recently, cryptic clade V was formally recognized as a separate species (*E. marmotae*), leaving five cryptic clades that have not been delineated at the species level. Here we report the novel species *Escherichia ruysiae* sp. nov., isolated from faecal material of an international traveller. *Escherichia ruysiae* sp. nov. encompasses the closely related *Escherichia* cryptic clades III and IV.

Isolation and Ecology

We discovered a cryptic clade IV strain in our collection, previously identified as extended spectrum beta-lactamase (ESBL) producing *E. coli* as part of the COMBAT study, which investigated the acquisition of ESBL-producing Enterobacteriaceae (ESBL-E) during international travel (15). This isolate, OPT1704^T, was further characterized in detail.

The strain was isolated from a human faecal sample provided immediately after an individual's return from a one-month journey to several Asian countries. No ESBL-E were detected in a faecal sample collected immediately before departure, suggesting the ESBL gene, and possibly strain OPT1704^T, were acquired during travel. The traveller reported diarrhoea during travel but no antibiotic usage. No ESBL-E were isolated in follow-up faecal samples, suggesting loss of the OPT1704^T strain or the ESBL gene within one month after return from travel.

Genome Features

The whole-genome sequence of strain OPT1704^T was determined using a combination of the Illumina HiSeq and Oxford Nanopore Technologies (ONT) sequencing platforms. The Illumina sequencing run yielded a total of 6.3×10^6 paired-end reads, with a mean read length of 151 bp. Illumina reads were downsampled using seqtk (version 1.3-r106, https://github.com/lh3/seqtk) to provide a theoretical coverage depth of 100X with the assumption that the OPT1704^T has a genome size of approximately 5×10^6 bp. The ONT sequencing run yielded a total of 2.5×10^4 reads, with a mean read length of 9078 bp before filtering. ONT reads were filtered on length and on read identity using Filtlong (version 0.2.0, https://github.com/rrwick/Filtlong) with Illumina reads as a reference, leaving 1.5×10^4 reads with a mean length of 12580 bp. This provided a theoretical coverage depth of

~38X of ONT reads. The combined assembly using Unicycler (version 0.4.6 (16)) of Illumina and Nanopore reads resulted in a completely assembled genome, consisting of one circular chromosome and one circular plasmid. The GC content of the complete OPT1704^T genome was 50.6%.

Putative resistance and virulence genes were predicted from the draft genome using ABRicate with the CARD (17) and VFDB (18) databases. OPT1704^T harbours 6 resistance genes on its IncI plasmid, associated with reduced susceptibility to fluoroquinolones (*qnrS1*), aminoglycosides (*aph(6)-Id* & *aph(3'')-Ib*), cephalosporins (*blaCTX-M-14*), trimethoprim (*dfrA14*) and sulphonamides (*sul2*), corresponding with its reduced susceptibility to fluoroquinolones (norfloxacin, MIC: 2 mg/L and ciprofloxacin, MIC: 0.5 mg/L), cephalosporins (cefuroxime, MIC: >32 mg/L and cefotaxime, MIC: 4 mg/L) and trimethoprim-sulfamethoxazole (MIC: >8 mg/L), assessed using VITEK2 (BioMérieux). However, strain OPT1704T was susceptible to tobramycin (MIC: ≤1 mg/L) and gentamicin (MIC: ≤1 mg/L) despite presence of aminoglycoside resistance genes. Furthermore, several putative virulence genes were predicted from the genome sequence associated with siderophore function (*chuX*, *entS*, *fepABD*), fimbriae (*fimBCDGI*), a type II secretion system (*gspGHI*) and capsular polysaccharide biogenesis (*kpsD*). These predicted virulence genes, when present in *Escherichia coli,* are not typically associated with a specific clinical syndrome such as diarrhoeal disease.

Physiology and Chemotaxonomy

Strain OPT1704^T formed circular, grey-white colonies on a Columbia sheep (COS) blood agar plate. Individual cells were observed under a light microscope and were rod-shaped and approximately 1 by 2 µm in size. The strain was shown to be Gram-negative, non-motile, oxidase-negative and catalase-positive. The strain was capable to grow in the absence of oxygen. On COS blood plates, it showed growth in the temperature range of 20-42 °C. The strain was also able to grow in NaCl concentrations ranging from 0% to 6% in lysogeny broth. MALDI-TOF (Bruker) and VITEK2 (BioMérieux) systems both identified OPT1704^T as *E. coli* with high confidence scores (score>2 for

MALDI-TOF and "Excellent identification" for VITEK2). Comparison of the output of the VITEK2 biochemical test with published biochemical reactions of other *Escherichia* species revealed that *E. ruysiae* sp. nov. str. OPT1704^T is distinct from other *Escherichia* species based on a biochemical profile (table 1) (2–4,19). One biochemical reaction, lysine decarboxylation, cannot be performed by strain OPT1704^T, but can be performed by all other *Escherichia* species. This reaction is typically mediated by the cadA gene (20), which is missing in OPT1704^T but present in other *Escherichia*.

16S rRNA and whole-genome phylogeny

Next, we calculated 16S rRNA sequence similarities, ANI values and digital DNA:DNA hybridisation (dDDH) values between OPT1704^T and type strains of the four other *Escherichia* species, representative genomes of the other three *Escherichia* cryptic clades, and *S. enterica* serovar Typhimurium (table 2). Representative genomes for the *Escherichia* cryptic clades I, II, III and VI were selected from Enterobase (21), using the genomes with the highest contiguity. Clades VII and VIII in Enterobase only consisted of a single strain and were not used in further analyses. We used three separate tools to calculate average nucleotide identity (ANI) (fastANI (22), OrthoANIu (23) and ANI calculator from Enveomics (24)) and the DSMZ Genome-to-Genome Distance Calculator to calculate digital DNA:DNA hybrisation (dDDH) (25). 16S rRNA genes were extracted from whole genomes using barrnap (version 0.9, https://github.com/tseemann/barrnap) and similarity was assessed using snp-dists (version 0.6, https://github.com/tseemann/snp-dists).

OPT1704^T showed 98.7-98.9% 16S rRNA sequence similarity to *E. coli* ATCC 11775^T, *E. fergusonii* ATCC 35469^T and *E. marmotae* HT073016^T, which would not warrant assignment to a novel species based on the current threshold for species delineation (less than 98.7% sequence similarity). However, the threshold for species delineation on the basis of 16S rRNA sequence has changed often and thresholds of up to 99% sequence similarity have been proposed previously (Kim 2014). In

120 contrast, ANI analysis and dDDH did support assignment of OPT1704^T to a novel species, together
121 with the representative strain of *Escherichia* cryptic clade III (table 2). The analyses also confirmed
122 that OPT1704^T falls within the *Escherichia* genus. This novel species, encompassing both *Escherichia*
123 cryptic clades III and IV, was assigned *E. ruysiae* sp. nov. with OPT1704^T as the proposed type strain.

124 To gain a better understanding of the *Escherichia* genus, we produced two phylogenies, based on
125 16S rRNA sequence (Fig. 1) and on an alignment of 682 core genes (Fig. 2). In short, rRNA genes were
126 predicted from whole genomes using barrnap (version 0.9, https://github.com/tseemann/barrnap)
127 and a tree was generated using FastTree (version 2.1.10 (26)). For the core gene alignment, genomes
128 were first annotated with Prokka (version 1.14.0 (27)) and a core gene alignment was produced
129 using Roary (version 3.12.0 (28)) and MAFFT (version 7.307 (29)). The phylogeny was inferred using a
130 generalised time reversible model using base frequencies from the SNP alignment and free rate
131 heterogeneity (GTR+F+R4 model) in IQ-tree (version 1.6.6 (30)), as advised by ModelFinder (31).
132 Phylogenies were rooted on the *Salmonella enterica* serovar Typhimurium str. LT2^T genome. Both
133 phylogenies showed that strain OPT1704^T clusters closely with the strain MOD1-EC7259 from
134 *Escherichia* cryptic clade III, and away from the current *Escherichia* species.

135 Chun et al. (32) proposed that strains with >95-96% genome-wide ANI between each other should
136 be assigned to the same species. If cryptic clades III and clade IV would share >95-96% ANI, this
137 would mean both clades should be assigned to the same novel species, *E. ruysiae*. To assess this for
138 a larger number of strains than the type strains presented in table 2, we downloaded all available
139 WGS from clade III and clade IV strains from Enterobase and compared ANI between all genomes
140 using fastANI (version 1.1 (22)). This analysis revealed that within 32 clade III genomes, the median
141 ANI is 98.6% (range: 97.7%-99.9%), while within 31 clade IV genomes, the median ANI is 98.9%
142 (range: 98.6%-99.9%). Between clade III and clade IV genomes, the median ANI is 96.6% (range
143 96.2%-96.8%). This suggests clades III and IV should be assigned to the same novel species, *E. ruysiae*
144 sp. nov.

Currently, no IJSEM guidelines exist for the delineation of subspecies based on genomic data. However, *E. ruysiae* could potentially be delineated further into two subspecies (representing the current clades III and IV, respectively) in the future, after a type strain for cryptic clade III has been identified.

Description of *E. ruysiae* sp. nov.

Escherichia ruysiae (ruy'si.ae N.L. fem. n. after Anna Charlotte Ruys, professor of microbiology at the University of Amsterdam from 1940 to 1969). Cells are Gram-negative, facultatively anaerobic, non-sporulating, non-motile rods with a size of approximately 1 by 2 µm. Colonies are circular, convex, grey-white and semi-transparent when grown overnight at 37 °C on COS sheep blood agar plates. The species is catalase-positive and oxidase-negative and grows at temperatures between 20 and 42 °C and NaCl concentrations between 0% and 6% w/v. In the VITEK2 GN biochemical test set it yields a positive result for Beta-Galactosidase, D-Glucose, D-Maltose, D-Mannitol, D-Mannose, D-Sorbitol, D-Trehalose, Saccharose/Sucrose, D-Tagatose, Gamma-Glutamyl-Transferase, Fermentation Glucose, Tyrosine Arylamidase, Succinate Alkalinisation, Alpha-Galactosidase, Ornithine Decarboxylase, Courmarate, Beta-Glucoronidase, 0/129 Resistance (Comp.Vibrio.) and Ellman and negative for Ala-Phe-Pro-Arylamidase, Adonitol, L-Pyrrolydonyl-Arylamidase, L-Arabitol, D-Cellobiose, H_2S Production, Beta-N-Acetyl Glucosaminidase, Glutamyl Arylamidase Pna, Beta-Glucosidase, Beta-Xylosidase, Beta-Alanine Arylamidase Pna, L-Proline Arylamidase, Lipase, Palatinose, Urease, Citrate (Sodium), Malonate, 5-Keto-D-Gluconate, L-Lactate Alkalinisation, Alpha-Glucosidase, Beta-N-Acetyl-Galactosaminidase, Phosphatase, Glycine Arylamidase, Lysine Decarboxylase, L-Histidine Assimilation, Glu-Gly-Arg-Arylamidase, L-Malate Assimilation and L-Lactate Assimilation.

168 The 16S rRNA sequence is deposited in ENA under accession LR745848. Raw Illumina and Nanopore
169 whole-genome sequencing data, as well as the complete genome assembly are deposited under
170 project PRJEB34275.

171

172 AUTHOR STATEMENTS

173 *Funding information*

174 The COMBAT study was funded by Netherlands Organization for Health, Research and Development
175 (ZonMw; 50-51700-98-120) and EU-H2020 programme (COMPARE, 643476).

176 *Acknowledgements*

177 The authors would like to thank Rob Weijts and Patricia Brinke for their help in phenotypic
178 characterization of type strain OPT1704^{T} of *Escherichia ruysiae* sp. nov. and Arie van der Ende for the
179 helpful discussions. We thank SURFsara (www.surfsara.nl) for the support in using the Lisa Compute
180 Cluster.

181 *Ethical statement*

182 Not required.

183 *Conflicts of interest*

184 None.

185

186 ABBREVIATIONS

187 ONT: Oxford Nanopore Technologies

188 MIC: Minimum Inhibitory Concentration

189 COS: Columbia agar + Sheep blood

190 dDDH: digital DNA:DNA hybridisation

191 ANI: Average Nucleotide Identity

192 SNP: Single Nucleotide Polymorphism

193

194 # REFERENCES

195 1. Castellani A, Chambers AJ. Manual of tropical medicine. William Wood; 1919.

196 2. Farmer JJ, Fanning GR, Davis BR, O'Hara CM, Riddle C, Hickman-Brenner FW, et al. Escherichia
197 fergusonii and Enterobacter taylorae, two new species of Enterobacteriaceae isolated from
198 clinical specimens. J Clin Microbiol. 1985 Jan 1;21(1):77.

199 3. Huys G, Cnockaert M, Janda JM, Swings J. Escherichia albertii sp. nov., a diarrhoeagenic species
200 isolated from stool specimens of Bangladeshi children. Int J Syst Evol Microbiol.
201 2003;53(3):807–10.

202 4. Liu S, Jin D, Lan R, Wang Y, Meng Q, Dai H, et al. Escherichia marmotae sp. nov., isolated from
203 faeces of Marmota himalayana. Int J Syst Evol Microbiol. 2015;65(7):2130–4.

204 5. Alnajar S, Gupta RS. Phylogenomics and comparative genomic studies delineate six main clades
205 within the family Enterobacteriaceae and support the reclassification of several polyphyletic
206 members of the family. Infect Genet Evol. 2017;54:108–127.

207 6. Priest FG, Barker M. Gram-negative bacteria associated with brewery yeasts: reclassification of
208 Obesumbacterium proteus biogroup 2 as Shimwellia pseudoproteus gen. nov., sp. nov., and
209 transfer of Escherichia blattae to Shimwellia blattae comb. nov. Int J Syst Evol Microbiol.
210 2010;60(4):828–833.

7. Tamura K, Sakazaki R, Kosako Y, Yoshizaki E. Leclercia adecarboxylata Gen. Nov., Comb. Nov., formerly known asEscherichia adecarboxylata. Curr Microbiol. 1986;13(4):179–184.

8. Hata H, Natori T, Mizuno T, Kanazawa I, Eldesouky I, Hayashi M, et al. Phylogenetics of family Enterobacteriaceae and proposal to reclassify Escherichia hermannii and Salmonella subterranea as Atlantibacter hermannii and Atlantibacter subterranea gen. nov., comb. nov. Microbiol Immunol. 2016;60(5):303–311.

9. Russo TA, Johnson JR. Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important endemic problem. Microbes Infect. 2003 Apr 1;5(5):449–56.

10. Savini V, Catavitello C, Talia M, Manna A, Pompetti F, Favaro M, et al. Multidrug-Resistant Escherichia fergusonii: a Case of Acute Cystitis. J Clin Microbiol. 2008 Apr 1;46(4):1551.

11. Ooka T, Seto K, Kawano K, Kobayashi H, Etoh Y, Ichihara S, et al. Clinical significance of Escherichia albertii. Emerg Infect Dis. 2012 Mar;18(3):488–92.

12. Liu S, Feng J, Pu J, Xu X, Lu S, Yang J, et al. Genomic and molecular characterisation of Escherichia marmotae from wild rodents in Qinghai-Tibet plateau as a potential pathogen. Sci Rep. 2019;9(1):1–9.

13. Gangiredla J, Mammel MK, Barnaba TJ, Tartera C, Gebru ST, Patel IR, et al. Draft Genome Sequences of Escherichia albertii, Escherichia fergusonii, and Strains Belonging to Six Cryptic Lineages of Escherichia spp. Genome Announc. 2018 May 3;6(18):e00271-18.

14. Walk S. The “Cryptic” Escherichia. EcoSal Plus [Internet]. 2015; Available from: https://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosalplus.ESP-0002-2015

15. Arcilla MS, van Hattem JM, Haverkate MR, Bootsma MCJ, van Genderen PJJ, Goorhuis A, et al. Import and spread of extended-spectrum β-lactamase-producing Enterobacteriaceae by international travellers (COMBAT study): a prospective, multicentre cohort study. Lancet Infect Dis. 2017 Jan 1;17(1):78–85.

16. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Comput Biol. 2017 Jun 8;13(6):e1005595.

17. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The Comprehensive Antibiotic Resistance Database. Antimicrob Agents Chemother. 2013 Jul;57(7):3348–57.

18. Chen L. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. 2004 Dec 17;33(Database issue):D325–8.

19. Abbott SL, O'Connor J, Robin T, Zimmer BL, Janda JM. Biochemical properties of a newly described Escherichia species, Escherichia albertii. J Clin Microbiol. 2003;41(10):4852–4854.

20. Snider J, Houry WA. MoxR AAA+ ATPases: a novel family of molecular chaperones? J Struct Biol. 2006;156(1):200–209.

21. Zhou Z, Alikhan N-F, Mohamed K, Fan Y, Achtman M, Brown D, et al. The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. Genome Res. 2020;30(1):138–152.

22. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018 Nov 30;9(1):5114.

23. Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek. 2017 Oct;110(10):1281–6.

24. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Prepr. 2016 Mar;4:e1900v1.

25. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 2013 Feb;14(1):60.

26. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS One. 2010;5(3).

27. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014 Jul 15;30(14):2068–9.

28. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015 Nov 15;31(22):3691–3.

29. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–780.

267 30. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic
268 Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2014 Nov
269 3;32(1):268–74.

270 31. Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermiin LS. ModelFinder: fast model
271 selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587.

272 32. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, et al. Proposed minimal
273 standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol
274 Microbiol. 2018 Jan 1;68(1):461–6.

275

276

FIGURES AND TABLES

Table 1. Comparison of biochemical markers which differentiate *E. ruysiae* sp. nov. from other *Escherichia* species. + and – indicate that ≥85% of tested strains is positive or negative for that biochemical marker, respectively. Data for *E. albertii, E. coli, E. fergusonii* and *E. marmotae* summarised from literature (Abbott 2003, Huys 2003, Farmer 1985, Liu 2015).

	E. ruysiae	*E. albertii*	*E. coli*	*E. fergusonii*	*E. marmotae*
ONPG	+	+	+	+	□–□
Lysine decarboxylase	-	+	+	+	+
Ornithine decarboxylase	+	+	+*	+	□–□
Fermentation of:					
Adonitol	-	□–□	□–□	+	□–□
d-Xylose	-	□–□	+	+	+
Cellobiose	-	□–□	□–□	+	□–□
d-Sorbitol	+	□–□	+	□–□	+

*50-85% of *E. coli* possess this biochemical property.

284 **Table 2.** Comparison of OPT1704^T 16S rRNA and whole-genome sequence with type strains of *E.*

285 *albertii*, *E. coli*, *E. fergusonii*, *E. marmotae*, representative genomes of *Escherichia* cryptic clades I, II,

286 III and VI and *S. enterica* serovar Typhimurium. In bold are the values that warrant assignment of

287 OPT1704^T to a novel species (<98.7% 16S rRNA sequence similarity, <95-96% ANI, <70% dDDH). ANI:

288 average nucleotide identity, dDDH: digital DNA:DNA hybridisation.

	E. ruysiae sp. nov. OPT1704^T				
	16S rRNA sequence similarity (%)	ANI (%, fastANI)	ANI (%, OrthoANIu)	ANI (%, ANI calculator Enveomics)	dDDH (%)
E. albertii NBRC 107761^T	**98.6**	**90.0**	**90.0**	**89.2**	**39.8**
E. coli ATCC 11775^T	98.7	**92.8**	**92.4**	**92.0**	**48.3**
E. fergusonii ATCC 35469^T	98.9	**89.4**	**88.2**	**89.7**	**36.7**
E. marmotae HT073016^T	98.9	**92.2**	**92.2**	**91.4**	**47.1**
S. enterica Typhimurium LT2^T	**97.5**	**82.1**	**80.7**	**81.8**	**24.0**
Escherichia cryptic clade I 89-3506	99.0	**92.5**	**92.1**	**91.8**	**47.8**
Escherichia cryptic clade II MOD1-EC7253	99.2	**92.0**	**91.7**	**91.0**	**45.5**
Escherichia cryptic clade III E4694	99.7	96.6	96.5	96.3	70.8
Escherichia cryptic clade VI UHCL_3L	**98.4**	**91.6**	**91.7**	**91.3**	**45.9**

289

290 **Figure 1**. 16S phylogeny of *E. ruysiae* str. OPT1704^T with type strains of other *Escherichia spp.*, other

291 *Escherichia* cryptic clades and *Salmonella enterica* serovar Typhimurium as outgroup. Numbers

292 indicate bootstraps on a scale of 0 to 1. Phylogeny available at

293 https://itol.embl.de/tree/145117226111975615791896 97

294

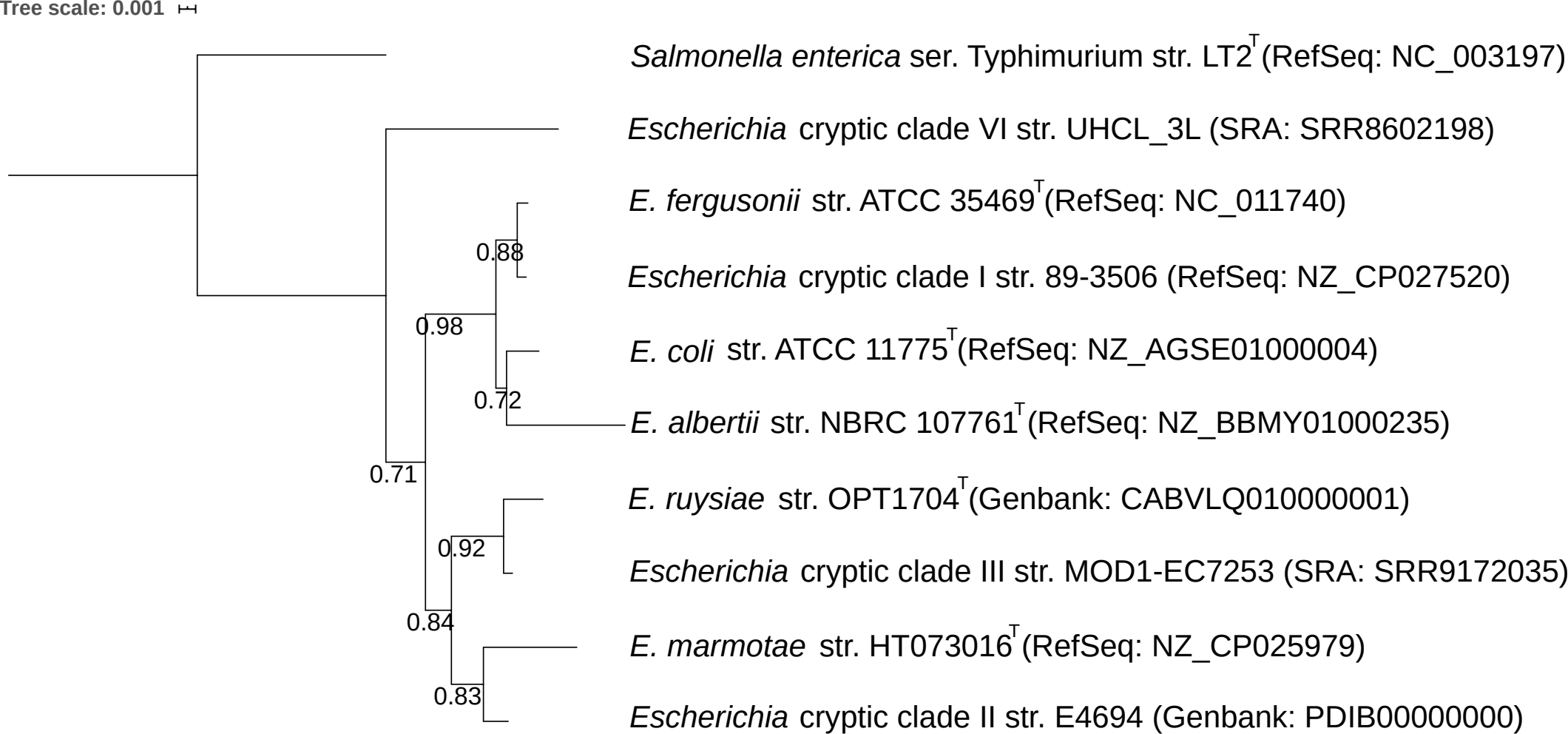
295

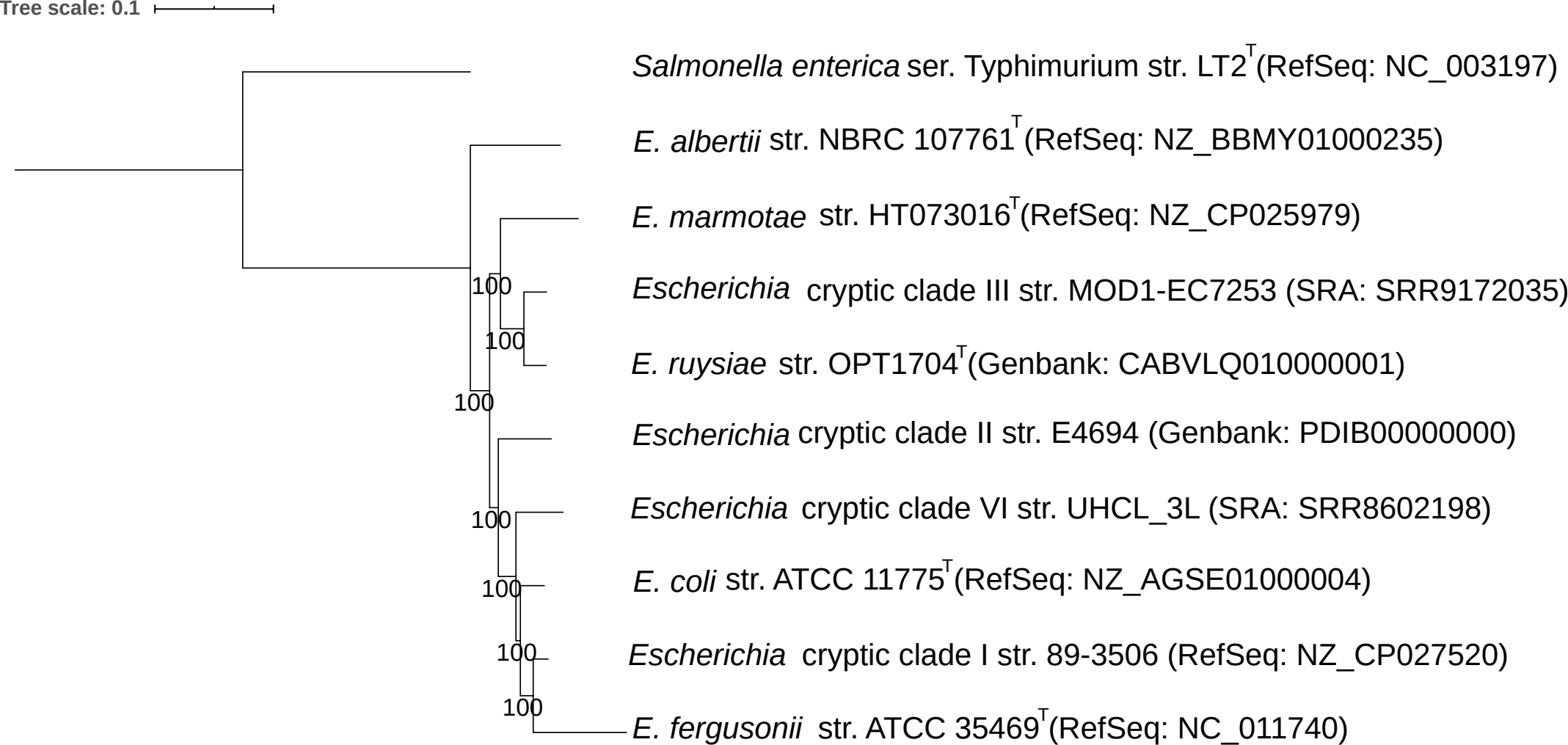
296 **Figure 2**. Phylogeny based on 682 concatenated core genes including *E. ruysiae* OPT1704T with type

297 strains of other *Escherichia spp.*, other *Escherichia* cryptic clades and *Salmonella enterica* serovar

298 Typhimurium as outgroup. Numbers indicate bootstraps on a scale of 0 to 100. Phylogeny available

299 at https://itol.embl.de/tree/14511722711358151579253758


Tree scale: 0.001
0.88
0.98
0.72
0.71
0.92
0.84
0.83
Salmonella enterica ser. Typhimurium str. LT2T (RefSeq: NC_003197)
Escherichia cryptic clade VI str. UHCL_3L (SRA: SRR8602198)
E. fergusonii str. ATCC 35469T (RefSeq: NC_011740)
Escherichia cryptic clade I str. 89-3506 (RefSeq: NZ_CP027520)
E. coli str. ATCC 11775T (RefSeq: NZ_AGSE01000004)
E. albertii str. NBRC 107761T (RefSeq: NZ_BBMY01000235)
E. ruysiae str. OPT1704T (Genbank: CABVLQ010000001)
Escherichia cryptic clade III str. MOD1-EC7253 (SRA: SRR9172035)
E. marmotae str. HT073016T (RefSeq: NZ_CP025979)
Escherichia cryptic clade II str. E4694 (Genbank: PDIB00000000)


Tree scale: 0.1
Salmonella enterica ser. Typhimurium str. LT2T (RefSeq: NC_003197)
E. albertii str. NBRC 107761T (RefSeq: NZ_BBMY01000235)
E. marmotae str. HT073016T (RefSeq: NZ_CP025979)
Escherichia cryptic clade III str. MOD1-EC7253 (SRA: SRR9172035)
E. ruysiae str. OPT1704T (Genbank: CABVLQ010000001)
Escherichia cryptic clade II str. E4694 (Genbank: PDIB00000000)
Escherichia cryptic clade VI str. UHCL_3L (SRA: SRR8602198)
E. coli str. ATCC 11775T (RefSeq: NZ_AGSE01000004)
Escherichia cryptic clade I str. 89-3506 (RefSeq: NZ_CP027520)
E. fergusonii str. ATCC 35469T (RefSeq: NC_011740)
100
100
100
100
100
100
100