THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Phylogenetic relationship and virulence composition of Escherichia coli O26:H11 cattle and human strain collections in Scotland; 2002-2020

Citation for published version:
Hoyle, D, Wee, BA, Macleod, K, Chase-Topping, M, Bease, A, Tongue, SC, Gally, D, Delannoy, S, Fach, P, Pearce, MC, Gunn, GJ, Holmes, A & Allison, L 2023, 'Phylogenetic relationship and virulence composition of Escherichia coli O26:H11 cattle and human strain collections in Scotland; 2002-2020', *Frontiers in Microbiology*, vol. 14, pp. 1-18. https://doi.org/10.3389/fmicb.2023.1260422

Digital Object Identifier (DOI):
10.3389/fmicb.2023.1260422

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Frontiers in Microbiology

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 13. Nov. 2023

1 # Phylogenetic relationship and virulence composition of *Escherichia coli*
2 # O26:H11 cattle and human strain collections in Scotland; 2002–2020

3

4 **AUTHORS**

5 **Deborah V. Hoyle[1]†[*], Bryan A. Wee[1]†, Kareen Macleod[1], Margo E. Chase-Topping[1], Andrew**
6 **Bease[1], Sue C. Tongue[2], David L. Gally[1], Sabine Delannoy[3], Patrick Fach[3], Michael C. Pearce[2],**
7 **George J. Gunn[2], Anne Holmes[4]††, Lesley Allison[4]††**

8

9 **AFFILIATIONS**

10 [1]Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush,
11 EH25 9RG.

12 [2]Centre for Epidemiology and Planetary Health, Department of Veterinary and Animal Science,
13 North Faculty, Scotland's Rural College (SRUC), Inverness, IV2 5NA, UK.

14 [3]Unité ColiPath – Plateforme IdentyPath, Laboratoire de Sécurité des Aliments, Agence Nationale
15 De Sécurité Sanitaire de l'alimentation, de l'environnement et du travail (ANSES), 14 rue Pierre et
16 Marie Curie, 94701 Maisons-Alfort Cedex, France.

17 [4]Scottish *E. coli* O157/STEC Reference Laboratory (SERL), Royal Infirmary of Edinburgh, 51 Little
18 France Crescent, EH16 4SA, UK.

19 †These authors share first authorship

20 ††These authors share last authorship

21

22 *Corresponding author: Deborah Hoyle, deborah.hoyle@ed.ac.uk

23

24 **Key words: STEC1, O26:H112, Cattle3, Phylogenetic4, whole genome sequencing (WGS)5,**
25 **Virulence6, Epidemiology7**

26

ABSTRACT

O26 is the commonest non-O157 Shiga toxin (*stx*)-producing *Escherichia coli* serogroup reported in human infections worldwide. Ruminants, particularly cattle, are the primary reservoir source for human infection. In this study we compared whole genomes and virulence profiles of O26:H11 strains (n=99) isolated from Scottish cattle with strains from human infections (n=96) held by the Scottish *Escherichia coli* O157/ STEC Reference Laboratory, isolated between 2002–2020. Bovine strains were from two national cross-sectional cattle surveys conducted between 2002–2004 and 2014–2015. A Maximum Likelihood phylogeny was constructed from a core genome alignment with the O26:H11 11368 reference genome. Genomes were screened against a panel of 2710 virulence genes using the Virulence Finder Database.

All *stx*-positive bovine O26:H11 strains belonged to the ST21 lineage and were grouped into three main clades. Bovine and human source strains were interspersed and *stx* subtype was relatively clade specific. Highly pathogenic *stx*2a-only ST21 strains were identified in two herds sampled in the second cattle survey and human clinical infections from 2010 onwards. The closest pairwise distance was 9 single nucleotide polymorphisms (SNPs) between Scottish bovine and human strains and 69 SNPs between the two cattle surveys. Bovine O26:H11 were compared to public EnteroBase ST29 complex genomes and found to have greatest commonality with O26:H11 strains from the rest of the UK, followed by France, Italy and Belgium.

Virulence profiles of *stx*-positive bovine and human strains were similar, but more conserved for the *stx2a* subtype. O26:H11 *stx*-negative ST29 (n=17) and ST396 strains (n=5) were isolated from 19 cattle herds; all were *eae*-positive and ten of these herds yielded strains positive for *ehxA*, *espK* and *Z2098,* gene markers suggestive of enterohaemorrhagic potential. There was a significant association ($p<0.001$) between nucleotide sequence percent identity and *stx* status for the bacteriophage insertion site genes *yecE* for *stx2*, and *yehV* for *stx1*. Acquired antimicrobial resistance genes were identified *in silico* in 12.1% of bovine and 17.7% of human O26:H11 strains, with s*ul2*, *tet*, *aph(3")* and *aph(6")* being most common. This study describes the diversity among Scottish bovine O26:H11 strains and investigates their relationship to human STEC infections.

DATA AVAILABILITY STATEMENT

The data sets for all raw genome sequence data analysed in this study can be found in the ENA (https://www.ebi.ac.uk/ena/browser/home) and SRA databases (https://www.ncbi.nlm.nih.gov/sra) under the accessions PRJEB57355 and PRJNA419720.

This is a provisional file, not the final typeset article

1. INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC) are a group of zoonotic pathogenic bacteria with a ruminant reservoir that cause gastrointestinal infections in humans [1,2]. Transmission to humans occurs via foodborne routes, as well as by direct contact with infected animals and through environmental contamination, particularly of water [3]. The Shiga toxin is the primary virulence factor responsible for severe pathology and is encoded by *stx* genes hosted on mobile lysogenic bacteriophage, which integrate into the bacterial genome at specific insertion sites [4]. There are two main Shiga toxin proteins, Stx1 and Stx2, encoded by different gene subtypes *stx1a, 1c, 1d, 1e* and *stx2a-o* [5–7], with the *stx2a*, *2c* and *2d* subtypes associated with more serious disease [8,9]. The majority of STEC are typically characterised by the presence of the locus of enterocyte effacement (LEE), which is required for the formation of attaching and effacing lesions in the intestine and encodes the intimin gene, *eae*, in addition to a number of other key virulence factors [10]. However, LEE is not essential for human pathogenicity, with some LEE-negative non-O157 STEC serotypes still capable of causing severe disease, mediated by other virulence determinants [11,12]. The LEE is not specific to the STEC pathotype and is also found in the majority of enteropathogenic *E. coli* (EPEC), which cause non-haemorrhagic gastro-intestinal illness in both animals and humans; EPEC are primarily distinguished from STEC by the absence of the *stx* gene [13].

STEC cause a spectrum of clinical symptoms in humans, from uncomplicated diarrhoea to haemorrhagic enteritis, haemolytic uraemic syndrome (HUS) and in exceptional cases, death. STEC serotypes that are responsible for the more severe, haemorrhagic disease presentations are further classified as enterohaemorrhagic (EHEC) [14]. Globally, *E. coli* O157:H7 is the commonest STEC/ EHEC serotype and is often associated with large foodborne outbreaks of disease. However, a number of non-O157 STEC serotypes can also be classed as EHEC based upon their disease and pathogenicity profile, with O26:H11 being the predominant non-O157 serotype of clinical relevance in human cases worldwide [15,16]. In Europe, O26:H11 now surpasses O157:H7 reported cases, and is currently the leading serotype responsible for human STEC infection, including paediatric HUS [17].

The recent increase in the proportion of STEC clinical cases attributed to non-O157 serotypes may in part be due to improvements in diagnostic testing methods [18]. However, the emergence of two highly pathogenic, *stx2*-only positive O26:H11 clones, termed the new "European" and "French" clones, has also resulted in a true increase in O26:H11 incidence across Europe over the past decade [19–21]. These clones have been particularly associated with disease outbreaks and hospitalisations in children, linked to the consumption of dairy produce in France, Italy and Romania [22–24].

Globally, O26:H11 strains can be grouped into two main multilocus sequence types (MLST), ST21 and ST29 [25]. The majority of all *stx*-positive O26:H11 belong to ST21, which includes the predominant *stx1*-only strains, as well as dual positive *stx1*+*stx2* strains and less common strains encoding *stx2a*, only. ST29 comprise mostly *stx*-negative O26:H11 strains, but also includes the newly emerging *stx2a*+ "European" and *stx2d*+ "French" clones. STEC harbour large virulence plasmids (pVF) that host genes for enterohaemolysin, *ehxA*, catalase peroxidase, *katP*, serine protease, *espP*, and a type II effector protein, *etpD* [19,26]. The main O26:H11 lineages are distinguished by the presence or absence of these pVF genes, with the ST21 lineage characterised by the *ehxA*+/*katP*+/*espP*+/*etpD*- gene profile. In contrast, the newly-emerging and highly virulent ST29 European *stx2*+ clone bears a distinct *ehxA*+/*katP*-/*espP*-/*etpD*+ pVF gene profile [19,25].

104 ST29 *stx*-negative O26:H11 strains that carry the *eae* gene are classed as EPEC, however a subset of
105 these strains have also been shown to carry the *ehxA*+/*katP*+/*espP*+/*etpD*- pVF gene profile typically
106 seen in *stx*-positive ST21 strains, together with a range of additional virulence factors [27]. Such *stx*-
107 negative strains have been termed "EHEC-like", because the acquisition of the *stx* gene through
108 bacteriophage lysogeny could result in conversion to a highly virulent EHEC pathogenic strain
109 profile [28,29]. In order to distinguish between O26:H11 *stx*-negative EPEC and EHEC-like strains,
110 an additional set of genetic markers has been proposed to assist in the identification of strains with
111 EHEC potential [29,30]. These markers include the type III secretion system genes *espK* [31], urease
112 gene, *ureD* [32] and the open reading frame putative marker Z2098 [33].

113 Scotland has a higher incidence of human STEC infections than the EU average and has reported an
114 increased incidence of non-O157 serotypes isolated from clinical patients in recent years [34,35]. In
115 order to assess the prevalence and distribution of STEC in Scottish cattle, two national cross-
116 sectional surveys were conducted in 2002–2004 [36,37] and 2014–2015 [38], from which a
117 collection of bovine sourced *stx*-positive and negative O26 *E. coli* strains were isolated [36,39]. The
118 aim of this present study was to compare by whole genome sequencing the O26:H11 strains isolated
119 from bovine faecal samples collected through these two Scottish cattle surveys with clinical O26:H11
120 human strains isolated from patients and previously sequenced by the Scottish *E. coli* O157/STEC
121 Reference Laboratory (SERL) [34]. We also further examined how the Scottish bovine strains related
122 to human-derived O26:H11 strains from across the wider UK [40,41] and investigated their global
123 O26:H11 phylogenetic context, by comparison with the public collection of clonal complex 29
124 genomes deposited within EnteroBase [42].

125 Ongoing analysis of strains from reservoir hosts such as cattle are essential for monitoring the
126 microevolution and emergence of new pathogenic STEC and EHEC strains. These data inform on
127 risk and can assist in the public health management of this pathogen.

128

This is a provisional file, not the final typeset article

2. MATERIALS and METHODS

2.1 Bacterial genomes included in the study

In total, 195 O26:H11 *E. coli* strains from Scottish cattle (n=99) and the Scottish human strain collection (n=96) were included in the analysis, together with three O177:H11 bovine strains that fell within the ST29 complex and a single bovine O103:H14 strain as an outgroup (Table 1; Supplementary Table 1).

Bovine strains were originally isolated from cattle faecal pat samples that had been obtained during two cross-sectional surveys of Scottish cattle farms conducted between 2002–2004 [36] and 2014–2015 [38,39], as previously described. In both surveys, the original faecal pat samples were collected by sampling discrete, dropped, faecal pats present on the ground of grazing land or the floor of pens.

The cattle strain collection comprised 60 isolates obtained from 35 herds, sampled in the 2002–2004 survey (Archive), and 43 isolates obtained from 29 herds in the 2014–2015 study (BECS). In the initial survey, Scotland was divided into six distinct geographical animal health district regions, as previously outlined [36] and herds were therefore also grouped according to this geographic classification in the second survey.

Human clinical O26:H11 genomes were provided from a collection of genome sequences held by the SERL. Clinical O26:H11 human strains were originally isolated from faecal sample submissions received by the SERL between 2002–2020 that were PCR-positive for *stx* genes. A subset of the human genomes sequenced at the SERL was included in this study, selected as described below in Phylogenetic analysis. Only a single representative genome from any outbreak-linked human strains was included in this comparative analysis.

Detailed methods used by the SERL for extraction, PCR, library preparation, sequencing and analysis have been described elsewhere [34]. In brief, genomic DNA was extracted either manually with the DNeasy Blood and Tissue Kit (Qiagen, Crawley, UK) or the QIAsymphony using the QIA DSP DNA Mini Kit (Qiagen). Libraries were prepared using the Nextera XT kit and sequenced on the Illumina MiSeq, producing paired-end reads of 250bp. Sequencing reads were processed in BioNumerics using the wgMLST and *E. coli* genotyping plug-in tools. The assembly was performed using SPAdes and basic assembly metrics were calculated for quality assessment. Sequencing reads were also processed using the Scottish Microbiology Reference Laboratory Edinburgh Bioinformatics Pipeline (SMiRLWBP). Trimmomatic [43] was used to removed bases with a Phred score < 30 from the trailing edge. KmerID [44] identified bacteria species and the GeneFinder tool mapped reads to a panel of serotype and virulence genes using Bowtie2 [45]. Only *in silico* predictions of serotype and virulence that matched a gene determinant at >80% nucleotide identity and over >80% target gene length were accepted. MLST alleles of seven housekeeping genes (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, *recA*) were determined using MOST (Metric-Oriented Sequence Typer) [46]. Shiga toxin gene subtyping was performed using a combined mapping and BLAST approach as previously described [47].

The funding bodies approved and authorised informed consent documentation for farm survey participants, to enable collection of dropped faecal pat samples from participant land. Permission had been granted and consent obtained for the samples, strains and data to be used for further research. All farm participant personal data were handled in accordance with the UK Data Protection Act (1998) and are now handled in accordance with the UK General Data Protection Regulation (GDPR 2018).

172 The de-identified bacterial genomes from human clinical samples were obtained from the National
173 Health Service (NHS) Lothian following ethical approval from the biorepository bank in NHS
174 Lothian covering the sequencing of samples (20/ES/0061).

175 **2.2 DNA extraction and sequencing of cattle-sourced bacterial strains.**

176 DNA was extracted from the cattle-sourced bacterial strains using the DNeasy Blood and Tissue Kit
177 (Qiagen, Crawley, UK) . The quantity of the DNA was measured using the Qubit Fluorimeter 3.0
178 (ThermoFisher Scientific) with the dsDNA Assay HS Kit. The *stx* subtype for all bovine isolates was
179 determined initially by PCR as previously described [36,39] and library preparation, sequencing and
180 analysis for serotype, MLST and Shiga toxin subtyping was performed at the SERL, as described
181 above.

182 **2.3 Phylogenetic analysis**

183 A subset of the Scottish human strain genomes held by the SERL that differed by fewer than 50
184 cgMLST (core genome MLST) alleles to bovine strains was determined using an ad-hoc cgMLST
185 schema with chewBBACA v2.5.5 [48], using allele profiles downloaded from EnteroBase
186 (https:/enterobase.warwick.ac.uk/schemes/Escherichia.cgMLSTv1) and made available for
187 comparison with the bovine genomes. This subset of human strain genomes were provided as paired
188 Illumina raw reads for full comparative analysis with all bovine strain genomes. Raw reads were
189 quality filtered using bbduk (v38.45) with the settings 'k=19 mink=11 hdist=1 ktrim=r
190 minoverlap=12 qtrim=rl trimq=20 minlength=50'[49]. All genomes were assembled using SPAdes
191 (v3.15.3) with the --careful option [50]. Criteria for inclusion were 50-51% GC, total assembly length
192 of 4.5 – 6.5Mbp, a duplication ratio below 1.021 and a minimum N50 of 40 Kbp. Genome alignment
193 was performed using Parsnp v1.5.6 with reference genome AP010953.1 (O26:H11 str. 11368) (4.03
194 Mbp, 35 612 polymorphic sites) [51]. Recombinant SNPs were filtered using Gubbins with default
195 settings (v3.0.0) [52]. A Maximum Likelihood tree (IQTree v2.1.2) was constructed from the filtered
196 core genome alignment of 198 sequences with the TVMe+ASC model and 1000 bootstraps [53].
197 Figures were generated using iTol [54].

198 **2.4 Virulence, phage insertion site and antimicrobial resistance gene identification**

199 All genomes were screened against a panel of 2710 virulence genes from the *E. coli* specific
200 virulence gene database (Date downloaded: 10 Sep 2021) using Abricate (v1.0.1), supplemented
201 with 139 additional gene alleles of interest, run with default parameters (--minid 80, --mincov 80)
202 (Supplementary Table 2; https://github.com/phac-nml/ecoli_vf) [55,56]. The percent identity was
203 recorded and gene target categorized as positive or negative, according to whether the gene was
204 detected using Abricate threshold parameters of minimum 80% coverage and 80% identity.
205 Additional targets included the *espK*, *ureD* and *Z2098* genes, thought to be indicative for potential
206 pathogenicity of O26:H11 *stx*-negative strains following acquisition of the *stx* gene [29,30].
207 Sequences for these genes were sourced from the O26:H11 reference strain 11368, GenBank
208 AP010953. Genomes were additionally screened using Abricate for the presence of four phage
209 insertion site genes commonly associated with *stx1* and *stx2* bacteriophage insertion into the O26
210 *E.coli* serogroup: *yecE* and *wrbA* for *stx2* and *yehV* and *sbcB* for *stx1* [4].

211 To examine the potential for antimicrobial resistance, strain genomes were screened for acquired
212 antimicrobial resistance genes (ARG) using StarAMR (v0.5.1), and the ResFinder gene database
213 (downloaded 07 Sep 2021) using default parameters (--pid-threshold 98, --percent-length-overlap 60)
214 [57,58]. Cattle strains positive for ARGs were submitted to the web-based Mobile Genetic Element

This is a provisional file, not the final typeset article

215 Finder (https://cge.food.dtu.dk/services/MobileElementFinder/) on 17th-30th October 2022, in order to
216 examine whether identified ARGs were associated with particular mobile genetic elements [59].

217 The local arrangement of genes within the genome was examined for the single integron bearing
218 bovine strain, and for exploration of bacteriophage insertion site genes in bovine strains, using
219 Artemis 18.1.0 [60]. Insertion site genes were identified using the navigator tool with primer
220 sequences according to Bonanno *et al.* 2015 [4]. Nucleotide and amino acid sequences were extracted
221 as fasta and verified in BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and Uniprot
222 (https://www.uniprot.org/) [61,62].

223 The potential carriage of any prophage sequences was assessed for all bovine *stx*-negative strains by
224 submission to PHASTER (https://phaster.ca/) in March/November 2022 [63].

225 **2.5 Comparison with publicly available worldwide genomes for *E. coli* ST29 complex**

226 Publicly available genomes from the *E. coli* ST29 complex (n=8511; Supplementary Table 3) were
227 downloaded from EnteroBase on 16 March 2022 [42]. The HeirCC HC-1100 cluster "2" was used to
228 filter genome assemblies and we further selected genomes for which "Country" metadata were
229 available (n=8332, from 36 countries).

230 **2.6 Pairwise distance in core genome alignment between Scottish bovine, human and publicly**
231 **available O26:H11 genomes**

232 A core genome alignment of the downloaded EnteroBase genomes in the *E.coli* ST29 complex,
233 together with 198 of the genomes from this study, was generated using Snippy (v4.6.0) against the
234 O26:H11 strain 11368 genome (AP010953.1) and the pairwise distances between genomes were
235 calculated using Disty (v0.1.0) [64,65]. A conservative threshold of 200 core SNPs was used to
236 capture clusters of epidemiologically linked isolates [66]. Distances were compared within herds,
237 between herds and between cattle and humans within Scotland, as well as between cattle to non-UK
238 country O26:H11 genomes.

239 **2.7 Statistical Analysis**

240 Fisher-Freeman-Halton Exact tests for comparing category proportions was performed in StatXact
241 Version 11 (Cytel Inc., Cambridge, MA, US). Associations between binary virulence gene
242 occurrence, *stx* profile and host species for all *stx*-positive O26:H11 genomes were performed using
243 nonmetric multidimensional scaling (NMS), PC-ORD software version 7.04 (MJM Software Design,
244 Gleneden Beach, OR, US). Strains that were *stx*-negative (n=22) were excluded from the analysis,
245 since the almost complete separation in virulence profiles between ST396 and the two differing
246 clusters of ST29 prevented the model from reaching a stable solution. Genes were excluded from the
247 analysis where they were present in <4% or >96% of the samples, identical to or highly correlated
248 with other genes (Supplementary Table 4; Supplementary Figure 1). The final NMS was run with 30
249 genes and 173 strain genomes. NMS was used with a grower distance measure. The dimensionality
250 of the data set was determined by plotting an inverse measure of fit ("stress") to the number of
251 dimensions. Optimal dimensionality was based on the number of dimensions with the lowest stress.
252 A three-dimensional solution was shown to be optimal. Several NMS runs were performed for each
253 analysis to ensure that the solution was stable and represented a configuration with the best possible
254 fit. On this basis, 500 iterations were used for each NMS run, using random starting coordinates.

255

256 **3. RESULTS AND DISCUSSION**

257 **3.1 Phylogenetic analysis of O26:H11 genomes**

258 All O26:H11 strains belonged to one of three sequence types: ST21 (87.2%, 170/195), ST29 (8.7%,
259 17/195) and ST396 (2.6%, 5/195), or other single locus variants of these STs (n=3) (Figure 1). All
260 *stx*-positive O26:H11 bovine strains belonged to ST21, as did the majority of the human strains
261 included in this analysis, whilst all *stx*-negative O26:H11 bovine strains belonged to either the ST29
262 or ST396 lineages. Three bovine O177:H11 strains that were also typed as ST29 were included in the
263 phylogeny: two *stx1a* positive strains from a single herd which also yielded a *stx*-negative O26:H11
264 strain and one *stx*-negative strain from a herd yielding *stx2a*+*stx1a* positive O26:H11 strains. This ST
265 distribution of O26:H11 strains in Scotland is highly similar to that reported for human cases in
266 England [38] and broadly in line with the most recent phylogenetic analysis of worldwide O26:H11
267 genomes by Long *et al.* [67], in which 84% and 16% of strains were identified as ST21 and ST29,
268 respectively.

269 Within the ST21 lineage, bovine and human *stx*-positive strains were interspersed within a relatively
270 diverse phylogeny, comprising three main clades (Figure 1). The first clade contained *stx2a*-only
271 positive O26:H11 cattle and human strains, which falls within the ST21C1a sub-lineage, as described
272 by Ogura and colleagues [25]. The second clade comprised predominantly *stx1a* positive strains only,
273 with two exceptions in human clinical strains, where *stx2a* was also present. The third clade
274 contained a combination of both *stx1*-only and dual *stx2*+*stx*1 positive strains. Strain ST and *stx*
275 subtype are outlined in Supplementary Table 1.

276 The first *stx2a*-only human strains isolated from clinical samples in Scotland were recorded in 2010
277 [34]. We did not identify any *stx2a*-only positive bovine strains in the 2002–2004 cattle survey
278 (n=338 sampled herds), however we isolated *stx2a*-only positive strains from two herds in the 2014–
279 2015 survey (n=110 sampled herds) that clustered with ST21 human *stx2a*-only strains (Table 1,
280 Supplementary Table 1). This *stx2a*-only lineage has not been reported in Scottish cattle prior to this
281 survey and are most likely explained by a relatively recent introduction into Scottish cattle, after the
282 estimated emergence of this clade in the mid-20th century [25].

283 We did not identify any Scottish *stx*-positive bovine strains belonging to the newly emerging ST29,
284 highly pathogenic *stx2a*-only new European clones [19,68] which has been isolated from clinical
285 cases throughout Europe and at low levels in Japan [69]. ST29 *stx2a*-only strains have been isolated
286 from <3% of Scottish clinical O26:H11 infections [34]. However, these strains were above the 50
287 cgMLST genetic distance threshold to Scottish bovine strains in our initial cgMLST-based screening,
288 the limit for selection of clinical strains for inclusion in the comparative phylogeny. Metadata
289 reported for clinical submissions to the national reference laboratory in England, indicated that ST29
290 *stx*2a-positive strains isolated from human cases reported within the UK were predominantly
291 associated with travel abroad [40]. This data would suggest that if ST29 *stx*2a-only O26:H11 are
292 present and circulating in the UK cattle population, this is not currently resulting in identifed human
293 infection.

294 We were interested in investigating the potential for O26:H11 pathogenicity in Scottish cattle
295 regardless of *stx* status and therefore sequenced strains from each of the nineteen herds yielding *stx*-
296 negative isolates (n=22). The majority of *stx*-negative strains belonged to ST29 (n=17 strains),
297 although four herds yielded minority strains (n=5) in ST396, a single-locus variant of ST29 [19]
298 (Figure 1). The ST29 genomes were split into two distinct clades, a relatively conserved clade

This is a provisional file, not the final typeset article

299 phylogenetically closer to the ST21 lineage and a more diverse clade at a greater distance. The
300 separation of *stx*-negative ST29 into distinct lineages, with one sublineage phylogenetically closer to
301 *stx*-positive ST21 is a broadly similar finding to that reported by Ogura *et al.* [25]. This also concurs
302 with the population structure observed in ST29 *stx*-negative O26:H11 strains isolated from both US
303 and New Zealand cattle populations [70,71] and with the recent clade classification of *stx*-negative
304 ST29 lineages by Long *et al*. [67]. Three Archive bovine strains which had been confirmed as *stx*-
305 positive by PCR at the SERL prior to long term cryostorage, were found to be *stx*-negative on
306 resuscitation, by both genome sequencing and repeat PCR. This may have been due to spontaneous
307 excision of the *stx*-encoding prophage.

308

309 **3.2 Pairwise distances between bovine isolates from the two Scottish cattle surveys**

310 Multiple O26:H11 isolates from individual herds were available for ten Archive and seven BECS
311 herds and 47 herds yielded a single strain per herd (Supplementary Table 1). Multiple strains within
312 herds were compared to examine within-herd diversity. Overall, the closest observed relationship was
313 within-herd/within-survey, with a minimum pairwise difference of 0 SNP and median of 3
314 (interquartile range, IQR = 5) (Figure 2, Supplementary Table 5), although strains within-herd from
315 the BECS survey had closer SNP distances than those within-herd from the Archive survey. Ten
316 *stx1a* genomes were available for a single Archive herd and gave a median pairwise difference of 4
317 SNPs and maximum of 15 SNPs, clustered within a single node. These data suggest that within the
318 study herds O26:H11 isolates with the same *stx*-profile spread clonally at the herd level, rather than
319 supporting multiple lineage introductions across a herd cohort. A similar observation has been
320 reported for O26:H11 within cattle herds in New Zealand [71]. However, where *stx* profile differed
321 within a herd, the SNP difference was found to be 128 SNPs or more, reflecting the presence of
322 different circulating lineages. Two herds yielded strains with the same *stx* subtype in both surveys,
323 however the genetic distance between these isolates was 69 SNPs or more. This also suggested the
324 presence of distinct lineages, since this pairwise distance is higher than the expected variation from a
325 single lineage over a 10-year period [40].

326 Regionally, across both cattle surveys, the median SNP difference between genomes from different
327 herds within an animal health district was 76 (IQR, 77) and between genomes across different animal
328 health districts the median SNP difference was 112 (IQR 63) (Figure 2, Supplementary Table 5).
329 These data indicate that herds that were geographically closer, showed closer genetic relationships
330 between strains. This is contrary to observations on regional differences reported for New Zealand
331 [71], and may be due to stock movements being more limited by geographical distance within
332 Scotland, particularly for herds based on Scottish islands.

333

334 **3.3 Pairwise distances between Scottish bovine isolates, Scottish human isolates and other**
335 **closely related O26:H11 genomes around the UK and globally**

336 A total of 3969 (47.6%) publicly available genomes within the HierCC:1100 (core genome ST
337 complex) clade 2 from EnteroBase, were found to be within 200 core SNPs from Scottish bovine strain
338 genomes (Supplementary Table 3, Supplementary Table 6). These genomes represented 24 countries;
339 countries excluded that fell beyond the 200 SNP threshold were from South America, Africa or Asia,
340 each representing fewer than five genomes, except for China, for which 11 genomes were available
341 (Supplementary Table 6).

342 The distribution profiles for pairwise SNP difference between bovine strains, between human strains
343 and between cattle to humans within Scotland and to humans across the rest of the UK were very
344 similar (Figure 3). All showed a biphasic distribution, due to the presence of two major clusters of
345 strains, with a similar median pairwise SNP difference in Scotland of cattle to cattle (between herds),
346 cattle to human, and human to human of 110 (IQR 63), 114 (IQR 66) and 118 (IQR 70), respectively
347 (Supplementary Table 5). The closest relationship between any bovine strain and a Scottish human
348 strain was 9 SNP between a BECS isolate (2014) and a human strain isolated in 2019. The closest
349 relationship to any UK human strain was 7 SNPs between this same BECS strain and a non-Scottish
350 UK human strain isolated from an individual with diarrhoea from the South of England in 2015
351 (Supplementary Table 6). For the latter example, the close relationship between the South England
352 human strain and the Scottish cattle strain could be due to either the movement of cattle or of a
353 bovine-contaminated food source, from Scotland to South England, or through human travel to and
354 consequent infection within Scotland. This study does not allow us to draw any conclusion on
355 directionality or source attribution, however these data could indicate a common source reservoir for
356 the majority ST21C1b lineage strains from cattle and humans within the UK.

357 The closest pairwise relationship observed between a Scottish bovine strain and an external country
358 genome was 10 SNP between three bovine strains from Archive_24 to a Canadian strain
359 (ESC_IB7316AA_AS), followed by 21 SNP between a 2019 French strain (ESC_FB8524AA_AS)
360 and a BECS_5 strain, and 26 SNP between an Archive_14 strain and a 2018 isolate from the United
361 States (ESC_RA4669AA_AS) (Figure 4, Supplementary Table 3, Supplementary Table 6). The strain
362 source for all these closest genomes was designated as "Human" origin in the EnteroBase "Source
363 Type" field. The 10 SNP difference between the Archive bovine strains and the human source
364 Canadian strain is unexpectedly close. The metadata associated with this Canadian strain indicates
365 source as a human with gastroenteritis, but does not provide information on isolation date. The
366 second closest Canadian strain at 33 SNPs to a BECS bovine strain is attributed to the Canadian Food
367 Inspection agency. The nearest pairwise difference to a designated bovine source genome in the
368 Canadian dataset was 134 SNPs, to a Canadian bovine faecal sample collected in 2014.

369 While two of the three non-UK strains that were closest matching to Scottish bovine strains were of
370 North American origin, overall, considering the proportion of genomes matched within 200 SNPs,
371 together with pairwise difference, Scottish cattle most closely matched strains from Europe than
372 elsewhere, with the greatest commonality seen with France and Belgium (Figure 4, Supplementary
373 Table 6). More than 60% of downloaded genomes from France, Italy, Belgium and Germany were
374 within 200 SNPs of our Scottish bovine strains, compared with less than 40% for North American
375 strains. Japan was an unusual outlier, matching 80% of genomes within 200 SNPs of Scottish cattle.
376 It has been previously noted that Japan imported relatively high levels of cattle from Western
377 countries during the second half of the 20th century [72], which may account for the closer
378 relationship of Scottish bovine strains to O26:H11 strains from Japan, than to other Asian-Pacific
379 countries. An analysis of source type by country was not performed due to a lack of available
380 metadata within EnteroBase for the majority of the downloaded genomes.

381 Defining a core genome alignment depends on the diversity and quality of genomes included in the
382 analysis. In this study, for the initial screen to identify Scottish human O26:H11 genomes that
383 clustered with the bovine O26:H11 genomes, we used a reference-free cgMLST clustering approach,
384 which would be less affected by sequence quality. For the subsequent in-depth analyses, we used the
385 more conservative Parsnp whole genome aligner to align Scottish human and bovine genomes. For
386 all comparisons of pairwise distances we used the more robust short sequence mapping-based
387 Snippy-Core-SNP approach, taken from the same core genome alignment generated using Snippy-

This is a provisional file, not the final typeset article

388 Core. The size of the alignment was approximately 276,190 SNPs. The pairwise distances calculated
389 here are only meaningful when used to compare subsets of genomes that were included in the same
390 core-genome alignment. This is because the core genome, by definition, is the collection of
391 nucleotide positions that are conserved across all the genomes in the given alignment and can change
392 according to the diversity of genomes being included.

393

394 **3.4 Virulence gene profiles**

395 All 195 O26:H11 genomes were screened against an *E. coli*-specific virulence gene database,
396 together with selected additional gene targets [56] and recorded as positive or negative according to
397 the described Abricate threshold parameters, (Supplementary Table 4, Supplementary Table 7). In
398 total, a conserved set of 154 genes, were identified as present in all genomes irrespective of *stx* status,
399 including the key virulence factor genes *eae*, *tir*, *cif*, *espA*, *espB*, encoded on the LEE pathogenicity
400 island. Other common virulence genes present across all strains included *fim D, fim F-H*, *gadX*, *iss*,
401 *lpfA*, and the non-locus of enterocyte effacement effector (*nle*) genes *nleB1*, *nleG7*, *nleG8* and *nleH1*.
402 This observation concurs with the typical O26:H11 virulence profiles previously reported in bovine
403 EPEC O26:H11 strains from the United States and in STEC O26:H11 worldwide [67,70].

404 A total of 207 genes displayed differential occurrence, with distinct distributions noted according to
405 *stx* status (Supplementary Table 7). All genomes were negative for the *etpD* gene, with 96.5% of all
406 *stx*-positive strains (n=167/173) showing the pVF profile typically observed in ST21 strains
407 *ehxA*+/*katP*+/*espP*+/*etpD*- (Figure 5). Three *stx*-positive Archive bovine strains from different herds
408 were *ehxA*-/*katP*-/*espP*-/*etpD*-, one Archive bovine and one human strain were *ehxA*+/*katP*+/*espP*-
409 /*etpD*-, with a further human strain having an *ehxA*+/*katP*-/*espP*-/*etpD*- pVF profile. The *stx*-negative
410 ST396 strains bore an identical core virulence profile to *stx*-positive strains, including the
411 *ehxA*+/*katP*+/*espP*+/*etpD*- profile. A further six *stx*-negative Scottish cattle herds yielded seven
412 strains that were *ehxA*+/*katP*-/*espP*-/*etpD*-, while the remaining nine *stx*-negative herds yielded ten
413 strains that were negative for all pVF genes. The *espL* and *fimB* genes showed similar distributions to
414 *ehxA* across all *stx* profiles, excluding the *stx*-negative ST396 lineage, of which all but one strain was
415 *fimB* negative. All but one of the *stx*-negative strains, regardless of ST or pVF profile, also carried a
416 distinct set of genes that were not detected in any *stx*-positive strains, including the genes *aec17,*
417 *aec18*, *aec22*, *aec*23, *hcp* and *vgrG,* which encode components of the type VI secretion system [73],
418 and the genes *Z0263* and *Z0265*.

419

420 **3.4.1 *stx-positive strain virulence profiles***

421 Other than *stx* subtype genes, the only gene specifically associated with *stx*-positive status was the
422 iron regulatory protein 1 gene, *irp1*, in 98.9% (n=171/173) of *stx*-positive strains, but not observed in
423 the *stx*-negative strains (Supplementary Table 7). In contrast, *irp2* was detected in all strains,
424 regardless of *stx* status. This is unusual, given that *irp1* and *irp2* are typically found together within a
425 high pathogenicity island. However, further analysis of a subset of these genomes using the Artemis
426 genome browser did identify an *irp1* variant allele in the *stx*-negative strains, bearing a nine base pair
427 insertion sequence, which presumably reduced the alignment to below the set Abricate threshold
428 parameters. The absence of *irp1* in the two *stx*-positive strains was found to be due to a contig break
429 within the gene. A further gene associated with *stx*-positivity was *nleG5-1*. This gene was found in
430 97.7% (n=169/173) of *stx*-positive strains, but only 9.1% (n=2/22) of *stx*-negative strains.

431 Strains that were positive for *stx2*-only were distinguished from all other *stx* profiles, including
432 negative strains, by the absence of the type VI secretory system gene *aec30*. However, other *aec*
433 subtypes such as *aec17-19, 22, 23* were either absent in all *stx* strains, but detected in the majority of
434 *stx*-negatives, or in the case of subtypes *aec24-29*, detected across all categories. Additional genes
435 that were absent in *stx2*-positive strains, but observed in all other *stx*-positive strains and in up to
436 50% of negative strains included the *espO1-1*, *iuc* and *iut* genes (Supplementary Table 7).

437 A non-metric multidimensional scaling (NMS) ordination model was constructed to examine
438 potential associations between binary virulence gene occurrence, *stx* profile and host species for all
439 *stx*-positive O26:H11 genomes (n=173) (Figure 6; Supplementary Table 8a). A three-dimensional
440 solution to the model was obtained which explained 83.5% of the variation (axis 1 = 40.1%, axis 2 =
441 27.3%, axis 3 = 16.1%). The graph was rotated to maximise the distance between cattle and human
442 strains on Axis 1 (Supplementary Figure 2). Axis 2 is explained by *stx1* (Kendall's tau, -4.28) and
443 *stx1*+*stx2* (Kendall's tau, 0.465) strains. *Stx2* was located primarily on axis 3 (Kendall's tau, 0.312)
444 (Figure 6). A multiple permutation test (MRPP) found significant differences between human and
445 bovine for *stx1* (bovine versus human, *p*=0.002) and *stx1*+*stx2* (bovine versus human, *p*=0.002).
446 There were no differences for *stx2*-only strains between bovine and humans (MRPP, *p*=0.158). Using
447 Kendall's tau as an indicator, most genes were not highly correlated with the NMS axes, though a
448 weak to moderate association was observed for axis 3 and the non-LEE effector genes *nleC* and
449 *nleG2-4* (Supplementary Table 8b).These data suggest that for our population, whilst some
450 differences exist in virulence background between the differing *stx* subtypes and host source within
451 the majority ST21 lineage, it was not possible to attribute this to specific genes. Virulence profiles
452 within the *stx2a*-only clade were more conserved than for the *stx1* and *stx1*+*stx2* strains, which
453 supports the phylogenetic analysis and observation that these strains have appeared in both Scottish
454 cattle and human strain populations only relatively recently.

455

456 **3.4.2 *stx-negative strain virulence profiles***

457 The 22 s*tx*-negative strains from 19 herds grouped into two core virulence profiles, "A" and "B",
458 across the ST29 and ST396 lineages, resulting in three distinct *stx*-negative populations overall
459 (Figure 5; Supplementary Table 9).

460 The 12 "A" profile strains, which included all ST396 strains from four herds and ST29 strains from
461 six herds, were located within the two clades phylogenetically closer to the *stx*-positive ST21 lineage.
462 The majority of "A" strains carried *ehxA*, *Z2201* and *espO1-1*, as well as the triplicate of genes, *ureD*,
463 *espK* and *Z2098*, suggested as key markers for identifying *E. coli* with the potential for EHEC-type
464 pathogenicity [30]. The presence of at least one of these three genes was always detected in all *stx*-
465 positive bovine and human strains. ST396, is a less common ST variant of ST29 and all strains in this
466 clade carried a greater complement of virulence genes, including the full *ehxA+*/*katP+*/*espP+*/*etpD*-
467 pVF profile, as well as in all except one strain, the three *espK*, *ureD* and *Z2098* genes. ST396 *stx2d*-
468 positive O26:H11 strains bearing the *ehxA+*/*kapP+*/*espP+*/*etpD*- profile have previously been
469 reported in a minority of human HUS cases from Italy [74].

470 The virulence gene profile borne by these "A" profile *stx*-negative strains, together with the
471 phylogenetic grouping, is consistent with an ST29C1 clade classification [25,67] and suggestive that
472 the strains are EHEC-like derivatives. The loss and acquisition of *stx* genes from O26 strains, both *in*
473 *vivo* and *in vitro*, has been previously documented [28,75]. Current diagnostic reliance on a positive

This is a provisional file, not the final typeset article

474 PCR result for *stx* and *eae* genes only, may therefore potentially result in false-negative classification
475 of EHEC strains that have lost the *stx* gene during laboratory isolation or within host. Our data
476 concur with the proposal by Delannoy *et al.* [30], that additional genes including *ehxA*, *espK*, *ureD*
477 and *Z2098* should be included in diagnostic screening assays and as shown here, are optimal gene
478 markers for the identification of O26:H11 EHEC potential in livestock and animal products.

479 A further nine herds yielded ten ST29 EPEC strains that were negative for *ehxA* and bore the
480 virulence profile "B". Distinguishing genes for this profile included *b2972*, the *gsp* cluster genes *C-*
481 *M*, *yghg*, *hlyA*, *lda(A-I)* and *ccdb* (Figure 5, Supplementary Table 9). This virulence profile is
482 consistent with the ST29C3 clade outlined by Long *et al.* [67] and observed elsewhere [27,29,70].
483 Strains in this clade, whilst bearing a combination of virulence factors found across varying *E. coli*
484 pathotypes [10], do not appear to have the appropriate virulence background for EHEC pathogenicity
485 following a potential recombination event with *stx*-bearing bacteriophage and are mostly represented
486 by EPEC strains. The *b2972* locus (*pppA* gene), *yghG* and *gsp(C-M)* are located in a common gene
487 cluster associated with the Type II secretion system found in both pathogenic and non-pathogenic *E.*
488 *coli* strains [76,77]. The *pppA* and *yghG* genes are also associated with regulation of heat-labile (LT)
489 toxin [78–80]. *hlyA* encodes α-haemolysin, an important cytotoxin found in uropathogenic *E. coli*
490 [81], while *ccdb* encodes a cytotoxin present in the toxin-antitoxin system targeting *E. coli* gyrase
491 and is carried by the F plasmid [82]. The *lda* genes, present in the locus of diffuse adherence, encode
492 adhesins associated with atypical EPEC and have also been reported in an O26:H11 paediatric
493 clinical strain [83].

494 One herd (BECS_2) yielded both ST21 *stx1a+stx2a* positive and ST396 *stx*-negative strains bearing
495 very similar key virulence genes, including *ehxA+*/*katP+*/*espP+*/*etpD-*, though missing *nleG-3*,
496 *nleG2-4*, *nleG5-1* and *fimB*. This herd also yielded an ST29 O177:H11 *stx*-negative strain, which
497 while *ehxA* positive, was *katP* and *espP* negative. In contrast, a second herd (BECS_28) yielded both
498 *stx1a* and *stx*-negative O26:H11 strains bearing the "B" virulence profile. A third herd (BECS_19)
499 generated an ST29 O26:H11 group "A" *stx*-negative strain and O177:H11 *stx1a* strains, both of
500 similar virulence profiles, though locating to different clades within the phylogeny.

501

502 **3.5 Phage insertion site genes**

503 The *stx* gene is encoded by mobile bacteriophage, which integrate into the bacterial host genome at
504 particular chromosomal insertion sites [84,85]. Insertion occurs within or adjacent to the host
505 insertion site gene and typically causes disruption to the insertion site gene sequence. A number of
506 integration sites have been identified for O26:H11 STEC, including the *yecE*, *wrbA*, *yehV* and *sbcB*
507 genes [4]. In order to examine whether there was any evidence for insertion of *stx*-phage at these sites
508 across the collection of strains in this study, we compared the gene sequence identity obtained from
509 the Abricate output, with presence and absence of *stx,* and the *stx* subtype. We observed variability in
510 nucleotide sequence percentage identity in the *stx* bacteriophage insertion site genes *yecE* and *yehV*,
511 according to *stx* status (Figure 7; Supplementary Table 10). *yehV* is one of the main insertion sites in
512 O26:H11 for the *stx*1 bacteriophage and *yecE* for *stx2* [4]. For *yecE*, variation in nucleotide
513 percentage identity was significantly associated with the presence of the *stx2* gene ($p < .001$, test
514 statistic 196, degrees of freedom 3), with 97% (64/66) of *stx2*-positive genomes showing a 90.6%
515 identity, and 98% (126/129) of *stx1*-only and *stx*-negative genomes showing 100% identity to *yecE*.
516 In contrast, significant variation in percent identity for *yehV* appeared to be associated with *stx1*
517 status ($p < .001$, test statistic 90.8, degrees of freedom 6), with the majority of the ST21 *stx1*-positive

518 genomes, as well as a subset of *stx*-negative ST29, showing 94.13% identity to *yehV*. There was
519 limited variability in percent identity across the genomes to *wrbA* and *sbcB*: 99% of genomes showed
520 96.37% identity to *wrbA* (193/195) and 100% identity to *sbcB* (194/195).

521 The individual insertion site gene size, sequence and gene arrangements in the vicinity of the *yecE*
522 and *yehV* genes were examined for a subset of bovine strains across the different *stx* categories using
523 the Artemis genome viewer [60] (Supplementary Table 1). In the *stx*-negative and *stx1*-positive
524 genomes examined, *yecE* was 819 base pairs (bp) in size and was located within the consecutive gene
525 sequence *yecD*, *yecE*, *yecN*, *cmoA* and *cmoB*. In contrast, the four *stx2*-only positive bovine strains
526 contained a truncated 111 bp fragment directly located next to the *yecD* gene, the latter immediately
527 adjacent to a contig break. The truncated 111 bp fragment had 100% homology to bases 1-96 of the
528 full length *yecE.* A further 816 bp gene was located on an alternate contig at a different location
529 within the genome, adjacent to the *yecN*, *cmoA* and *cmoB* genes, in most case flanked by an integrase
530 gene. This 816 bp gene showed 0% homology to *yecE* between bases 1-70 and 99% homology
531 between bases 71-819. All 13 *stx1*+*stx2* strains examined also showed a truncated 816 bp *yecE* gene
532 adjacent to an integrase gene and seven of these additionally bore the 111 bp fragment. This suggests
533 the potential occurrence of an integration event at the *yecE* site, resulting in disruption of this gene in
534 the *stx2*-positive strains, which was not observed in the *stx1*-only and *stx*-negative strains.

535 We examined the *yehV* (*mlrA*), gene length and arrangement in 6 *stx1*-positive, 13 *stx1*+*stx2* and 22
536 *stx*-negative cattle O26:H11 genomes. All *stx1*-positive, 11/13 *stx1*+*stx2* and the 7 "A" profile ST29
537 *stx*-negative genomes showed <100% homology to *yehV* and contained a 648 bp gene of 94.13%
538 identity to *yehV*, flanked in all cases by *yehW* and the integrase *IntQ_1* or *IntQ_2* genes [86]. The
539 presence of the truncated *yehV* gene flanked by integrase *intQ* genes is highly suggestive of a phage
540 insertion event in these 7 "A" profile ST29 *stx*-negative strains. In contrast, 9/10 *stx*-negative ST29
541 profile "B" strains with 100% homology to *yehV* , as well as the three O177:H11 strains, contained a
542 full length 732 bp *yehV* (*mlrA*) gene, flanked by *yehW* and the sensory histidine kinase gene *ypdA-1*.
543 One "B" profile ST29 *stx*-negative strain contained two smaller gene fragments, and two *stx1*+*stx2*
544 strains bore the full-sized gene

545 The ST396 "A" profile *stx*-negative strains carried a greater complement of virulence genes than the
546 ST29 *stx*-negative strains. However, the *yehV* local gene arrangement in these ST396 strains did not
547 show evidence of phage integration or interruption, with an intact 732bp *yehV* gene. We examined
548 the gene arrangement in these strains at another potential bacteriophage insertion site, the *torS-T*
549 intergenic region [87], however this region was uninterrupted in all the ST396 strains.

550 All *stx*-negative cattle genomes were submitted to PHASTER for identification of any *stx*-prophage
551 regions, however the results were inconclusive (Supplementary Table 9). Both intact and incomplete
552 prophage regions with homology to *stx*-prophage as the first or second listed most common phage
553 were identified in the majority of the *stx*-negative strains. Due to the limitations of short read genome
554 assembly, long read sequencing would be required to confirm the presence of any inserted prophage
555 in *stx*-negative strains. However, given the distinct gene arrangement in the locality of *yehV,* together
556 with the virulence background present in these strains, it would seem probable that they had either
557 the potential for acquiring *stx*, or had previously been *stx*-positive and subsequently lost the *stx* gene
558 from an integrated prophage.

559

560 **3.6 Antimicrobial resistance profiles of bovine strains**

561 Acquired antimicrobial resistance genes (ARGs) were identified *in silico* for 12.1% (12/99) of the
562 O26:H11 bovine strains, in 15.6% of the cattle herds (10/64) (Table 2). The commonest genes
563 detected were *sul2* and *tet*, found in six herds, *aph(3")* and *aph(6")* in five herds, and bla_{TEM} and
564 *dfrA*, each detected in four herds. Three-quarters of resistant strains (9/12) were positive for more
565 than one ARG (median 4.5, range 1-5), with the commonest combination being *aph(3")*, *aph(6")*,
566 together with one or more further ARGs. Seventeen human strains (17.7%) carried ARGs (Table 2),
567 with the commonest being *aph(3")* and *aph(6")* in 13 strains, followed by *sul2* and *floR* in eight
568 strains. Multiple resistance was also common in human strains (median 4, range 1-9).

569 Strains carrying resistance were screened for the presence of mobile genetic elements
570 (Supplementary Table 11). One bovine Archive genome carried a clinical class 1 integron with a
571 typical cassette arrangement of 5'–*intI1*, *dfrA1*, *aadA1*, *qacEΔ1*, *sul1*–3' with an additional two
572 genes, *floR* and *sul2* found in close proximity on the same contig. Two multiply resistant human
573 strains also carried clinical class 1 integrons. Across all genomes, bla_{TEM} genes were found on
574 transposon Tn2 in six cases, and in cattle *dfrA1* was associated with the composite transposon
575 cn_4568_IS26 in two herds.

576 The resistome seen in the bovine isolates was very similar to that observed in the Scottish human
577 dataset, with the commonest ARGs conferring resistance to streptomycin and spectinomycin
578 aminoglycosides, sulphonamides, tetracyclines and beta-lactam agents such as ampicillin, the latter
579 class designated as critically important antimicrobials for human health by the World Health
580 Organisation [88]. The proportion of the bovine strains carrying ARGs is in line with that reported in
581 a collection of O26:H11 strains from home (non-travel associated) human STEC O26 cases from
582 England & Wales isolated during 2015 [89], but slightly lower than described in a more recent report
583 on human case clonal complex 29 STEC isolates in England between 2014-2021 [41]. Our results
584 differ markedly from the very high AMR prevalence reported in a collection of O26:H11 strains from
585 feedlot cattle in the United States [70], however this likely reflects differences in the management
586 systems and associated antimicrobial usage of the livestock systems between the two countries.

587 Overall, these data are in accordance with current antimicrobial usage observed within the bovine
588 sector in the UK, with beta-lactams, tetracycline and streptomycin being the most frequently
589 prescribed antimicrobials in both beef and dairy cattle [90,91]. Given ruminants are the primary
590 reservoir source for human infection within the UK and antimicrobial therapy is generally not
591 indicated in human STEC infection[92], this agreement in resistome profile between the Scottish
592 cattle and human genomes is not unexpected.

593

594 **4. CONCLUSION**

595 In conclusion, within the study herds all *stx*-positive cattle O26:H11 strains fell within the ST21
596 lineage and no ST29 *stx*-positive strains were identified. Bovine and clinical human strain genomes
597 were relatively well interspersed, with *stx* subtype generally clade specific. Highly pathogenic *stx2a*-
598 only ST21 were identified in two herds from the second cattle survey and in human strains from 2010
599 onwards. Where multiple strains were available from individual herds we observed limited variability
600 within *stx*-subtype, suggesting same *stx*-subtype strains typically spread clonally at the farm level,
601 rather than supporting multiple lineage introductions across a cohort. Half of the *stx*-negative survey
602 herds yielded O26:H11 strains with virulence profiles similar to that observed in *stx*-positive strains,
603 including the genes *ehxA*, *espK*, *Z2098*, which have been proposed as markers for "EHEC-like"

604 potential. These data suggest that the reservoir of O26:H11 in Scottish cattle bearing a genomic
605 background compatible with EHEC potential and therefore of public health concern may be greater
606 than would be expected based on detection of the STEC markers *stx* and *eae* alone.

607

608

This is a provisional file, not the final typeset article

609 **TABLES**

610 **Table 1.** Summary of the bovine and human bacterial strain genomes included in this study.

Strain Collection (Year)	**Serotype**	***stx* gene profile**	**Number of genomes**	**Reference**
Archive cattle (2002–2004)	**O26:H11**	*stx1*	41	Pearce *et al.* 2006 [36]
		stx1+*stx2*	16	
		stx-negative	3	
BECS cattle (2014–2015)	**O26:H11**	*stx1*	10	Hoyle *et al.* 2021 [39]
		stx2	4	
		stx1+*stx2*	6	
		stx-negative	19	
	O177:H11	*stx1*	2	
		stx-negative	1	
	O103:H14	*stx*-negative	1	
Human (SERL) (2002–2020)	**O26:H11**	*stx1*	56	Food Standards Scotland, 2020 [34]
		stx2	6	
		stx1+*stx2*	34	

611

612

613 **Table 2.** Summary of antimicrobial resistance genes detected in the Scottish bovine and human
614 O26:H11 genomes by STARAMR and ResFinder databases, where P indicates genome positive for
615 respective genes.

Strain ID	**Source**	***ant(3")* or *aadA1***	***aph(3")-Ib***	***aph(6")-Id***	***bla*$_{OXA-1}$**	***bla*$_{TEM-1B}$**	***bla*$_{TEM-1C}$**	***bla*$_{TEM-30}$**	***dfrA1***	***dfrA5***	***floR***	***mph(B)***	***sul1***	***sul2***	***tet A***	***tet B***	***tet C***
XH2001256	Archive_31															P	
XH2001264r	Archive_14	P							P		P		P	P			
XH2001404	Archive_35		P	P				P						P	P		
XH800939X	BECS_2		P	P			P		P					P			
XH800941P	BECS_2		P	P			P		P					P			
XH800951Y	BECS_16		P	P			P		P					P			
XH800956T	BECS_25									P					P		
XH800958M	BECS_27		P	P		P								P	P		
XH800985H	BECS_12		P	P							P			P			
XH800986Y	BECS_12		P	P							P			P			
XH800989A	BECS_24																P
XH801004W	BECS_6																P
SME-18-85	Human					P											
SME-18-152	Human		P	P		P								P			
SME-18-190	Human	P	P	P					P		P	P	P	P	P		
SME-18-45	Human		P	P													
SME-18-27	Human		P	P		P					P			P	P		
SME-18-201	Human		P	P		P					P			P	P		
SME-18-138	Human	P											P		P		
SME-18-30	Human		P	P							P			P			
SME-18-194	Human		P	P			P							P			
SME-18-88	Human		P	P							P			P			
SME-18-10	Human		P	P				P									
MUOON6	Human		P	P													
SME-19-228	Human		P	P							P						
SME-18-195	Human		P	P													
SME-19-812	Human													P			
SME-20-404	Human	P			P						P			P			
SME-20-481	Human		P	P							P			P			

616

This is a provisional file, not the final typeset article

617 **FIGURE LEGENDS**

618 **Figure 1.** A Maximum Likelihood core genome phylogenetic analysis of Scottish bovine and human
619 O26:H11 isolates (n=195) and three O177:H11 strains (*) analysed in this study. The inner ring
620 shows the presence of the *stx1* and *stx2* genes and the outer ring shows the MLST number. The tips
621 are coloured by source and dataset (cattle – BECS [orange] or Archive [yellow]; human clinical
622 [grey]). Red circles indicate branches with >90% bootstrap support. Tree scale is in substitutions per
623 site.

624 **Figure 2.** Histogram showing pairwise SNP differences <200 SNP, between bovine O26:H11 strains
625 (n=99), for the comparisons "Within Herd", "Within Region" (excludes Within Herd), "Between
626 Regions" and "Between All Herds".

627 **Figure 3.** Histogram showing pairwise SNP differences between bovine O26:H11 strains (n=99),
628 Scottish human (n=96) and rest of UK human (n=1217), displaying comparison with <200 SNP
629 difference. Comparisons given as "Cattle to Scottish Human", "Between Scottish Cattle Herds",
630 Cattle to rest UK Human" and "Between Scottish Human". Pairwise comparisons include bovine
631 strains across both surveys.

632 **Figure 4.** Raincloud plot for the pairwise SNP difference between bovine O26:H11 strains (n=99) to
633 publicly available O26:H11 genomes from other countries present within the EnteroBase *E.coli* ST29
634 complex. Pairwise comparisons displayed for bovine strains across both surveys to countries where
635 five or more genomes match to bovine strains < 200 SNPs. Error bars illustrate the median and
636 interquartile range.

637 **Figure 5.** A Maximum Likelihood core genome phylogenetic analysis of Scottish bovine and human
638 O26:H11 isolates (n=195) and three O177:H11 strains (*), showing the presence of 14 selected key
639 virulence genes or gene clusters. From inner to outer ring (*ehxA*, *katP*, *espP*, *ureD*, *espK*, *Z2098*, *iuc-*
640 *iut*, *espO1-1*, *aec30*, *aec17*, *hlyA*, *gsp-yghg*, *lda*, *ccdb*). Red circles indicate branches with >90%
641 bootstrap support. Tree scale is in substitutions per site.

642 **Figure 6.** Non-metric multidimensional scaling ordination model to examine potential associations
643 between a subset of differential gene occurrence for host source-*stx* profile combined (Human,
644 Bovine; *stx1*: stx1, *stx2*: stx2, *stx1*+*stx2*: stx12), for all *stx*-positive genomes (n=173). Arrows
645 indicate vector direction.

646 **Figure 7.** A Maximum Likelihood core genome phylogenetic analysis of Scottish bovine and human
647 O26:H11 isolates (n=195) and three O177:H11 (*) showing the sequence conservation of four known
648 phage insertion sites (*yecE*, *yehV*, *sbcB*, *wrbA*). Red circles indicate branches with >90% bootstrap
649 support. Tree scale is in substitutions per site.

650

651

652

653

654

655 **CONFLICTS OF INTEREST**

656 The authors declare that the research was conducted in the absence of any commercial or financial
657 relationships that could be construed as a potential conflict of interest.

658 **AUTHOR CONTRIBUTIONS**

659 D.V.H., A.H., L.A. and D.G., conceptualized the study. D.V.H. obtained funding for the study.
660 D.V.H. and B.W. analysed all the data and wrote the original manuscript. A.H. and L.A. sequenced
661 the isolates and curated data. B.W. performed bioinformatics analyses, including genome assembly,
662 polishing, annotation, phylogenetics and tree diagram generation. D.V.H carried out Artemis strain
663 analysis. K.M. isolated bovine strains from the BECS survey and prepared DNA. S.C.T., M.C.P. and
664 G.J.G. performed the original cattle surveys and provided samples from which bovine isolates were
665 generated. M.C.T performed statistical analyses. A.B. performed data curation. S.D. and P.F.
666 provided guidance for interpretation of O26:H11 genomes. All listed authors reviewed and edited the
667 manuscript.

668 **FUNDING INFORMATION**

669 This research was funded by the Wellcome Trust, through a personal fellowship to D.V.H.
670 (105832/Z/14/Z). Additional resources to D.V.H and part funding for B.W. were provided through an
671 Institute Strategic Programme grant from the Biotechnology and Biological Sciences Research
672 Council (BBS/E/D/20002173). B.W. was also part funded by the Novo Nordisk Foundation Fund
673 (Grant: NNF16OC0021856: Global Surveillance of Antimicrobial Resistance).

674 **ACKNOWLEDGEMENTS**

675 We thank Sharif Shaaban for assisting with the initial ChewBBACA comparison of clinical and cattle
676 genomes and Stella Mazeri for assisting with R script provision for figure generation. We are grateful
677 to the staff of the Scottish O157/STEC Reference Laboratory for laboratory assistance and to all the
678 farmers who participated in the original Archive and BECS cattle surveys.

679 For the purpose of open access, the author has applied a CC-BY public copyright licence to any
680 Author Accepted Manuscript version arising from this submission.

681

REFERENCES

1 World Health Organization, Food and Agriculture Organization of the United Nations. Shiga toxin-producing *Escherichia coli* (STEC) and food: attribution, characterization, and monitoring. Published Online First: 2018.https://apps.who.int/iris/handle/10665/272871 (accessed 12 Dec 2022).

2 Kolenda R, Burdukiewicz M, Schierack P. A systematic review and meta-analysis of the epidemiology of pathogenic *Escherichia coli* of calves and the role of calves as reservoirs for human pathogenic *E. coli*. Front Cell Infect Microbiol. 2015;**5**:23. doi:10.3389/fcimb.2015.00023

3 Kintz E, Brainard J, Hooper L, *et al.* Transmission pathways for sporadic Shiga-toxin producing *E. coli* infections: A systematic review and meta-analysis. Int J Hyg Env. Health. 2017;**220**:57–67. doi:10.1016/j.ijheh.2016.10.011

4 Bonanno L, Loukiadis E, Mariani-Kurkdjian P, *et al.* Diversity of Shiga toxin-producing *Escherichia coli* (STEC) O26:H11 strains examined via *stx* subtypes and insertion sites of *stx* and *espk* bacteriophages. Appl Env. Microbiol. 2015;**81**:3712–21. doi:10.1128/AEM.00077-15

5 Scheutz F, Teel LD, Beutin L, *et al.* Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. J Clin Microbiol. 2012;**50**:2951–63. doi:10.1128/JCM.00860-12

6 Hughes AC, Zhang Y, Bai X, *et al.* Structural and functional characterization of Stx2k, a new subtype of Shiga toxin 2. *Microorganisms* 2019;**8**:4. doi:10.3390/microorganisms8010004

7 Gill A, Dussault F, McMahon T, *et al.* Characterization of Atypical Shiga Toxin Gene Sequences and Description of Stx2j, a New Subtype. *J Clin Microbiol* 2022;**60**:e0222921. doi:10.1128/jcm.02229-21

8 Friedrich AW, Bielaszewska M, Zhang W-L, *et al. Escherichia coli* harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. *J Infect Dis* 2002;**185**:74–84. doi:10.1086/338115

9 Persson S, Olsen KEP, Ethelberg S, *et al.* Subtyping method for *Escherichia coli* Shiga toxin (Verocytotoxin) 2 variants and correlations to clinical manifestations. *J Clin Microbiol* 2007;**45**:2020–4. doi:10.1128/JCM.02591-06

10 Kaper JB, Nataro JP, Mobley HLT. Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2004;**2**:123–40. doi:10.1038/nrmicro818

11 Newton HJ, Sloan J, Bulach DM, *et al.* Shiga Toxin–producing Escherichia coli Strains Negative for Locus of Enterocyte Effacement. *Emerg Infect Dis* 2009;**15**:372–80. doi:10.3201/eid1502.080631

12 Colello R, Krüger A, Velez MV, *et al.* Identification and detection of iha subtypes in LEE-negative Shiga toxin-producing Escherichia coli (STEC) strains isolated from humans, cattle and food. *Heliyon* 2019;**5**:e03015. doi:10.1016/j.heliyon.2019.e03015

718 13 Denamur E, Clermont O, Bonacorsi S, *et al.* The population genetics of pathogenic Escherichia
719 coli. *Nat Rev Microbiol* 2021;**19**:37–54. doi:10.1038/s41579-020-0416-x

720 14 Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* 1998;**11**:142–201.
721 doi:10.1128/CMR.11.1.142

722 15 Caprioli A, Morabito S, Brugère H, *et al.* Enterohaemorrhagic *Escherichia coli*: emerging issues
723 on virulence and modes of transmission. *Vet Res* 2005;**36**:289–311. doi:10.1051/vetres:2005002

724 16 Johnson KE, Thorpe CM, Sears CL. The emerging clinical importance of non-O157 Shiga toxin-
725 producing *Escherichia coli*. Clin Infect Dis. 2006;**43**:1587–95. doi:10.1086/509573

726 17 European Centre for Disease Prevention and Control. Shiga toxin-producing *Escherichia coli*
727 (STEC) infection - Annual epidemiological report for 2020. *Eur Cent Dis Prev Control*
728 Published Online First: 12 September 2022.https://www.ecdc.europa.eu/en/publications-
729 data/shiga-toxin-producing-escherichia-coli-stec-infection-annual-epidemiological-0 (accessed
730 12 Dec 2022).

731 18 Parsons BD, Zelyas N, Berenger BM, *et al.* Detection, characterization, and typing of Shiga
732 toxin-producing *Escherichia coli*. Front Microbiol. 2016;**7**:478. doi:10.3389/fmicb.2016.00478

733 19 Bielaszewska M, Mellmann A, Bletz S, *et al.* Enterohemorrhagic *Escherichia coli* O26:H11/H-: a
734 new virulent clone emerges in Europe. Clin Infect Dis. 2013;**56**:1373–81. doi:10.1093/cid/cit055

735 20 Zweifel C, Cernela N, Stephan R. Detection of the emerging Shiga toxin-producing *Escherichia*
736 *coli* O26:H11/H- sequence type 29 (ST29) clone in human patients and healthy cattle in
737 Switzerland. Appl Env. Microbiol. 2013;**79**:5411–3. doi:10.1128/AEM.01728-13

738 21 Delannoy S, Mariani-Kurkdjian P, Bonacorsi S, *et al.* Characteristics of emerging human-
739 pathogenic *Escherichia coli* O26:H11 strains isolated in France between 2010 and 2013 and
740 carrying the *stx2d* gene only. J Clin Microbiol. 2015;**53**:486–92. doi:10.1128/JCM.02290-14

741 22 Jones G, Lefevre S, Donguy MP, *et al.* Outbreak of Shiga toxin-producing *Escherichia coli*
742 (STEC) O26 paediatric haemolytic uraemic syndrome (HUS) cases associated with the
743 consumption of soft raw cow's milk cheeses, France, March to May 2019. Euro Surveill.
744 2019;**24**. doi:10.2807/1560-7917.ES.2019.24.22.1900305

745 23 Severi E, Vial F, Peron E, *et al.* Community-wide outbreaks of haemolytic uraemic syndrome
746 associated with Shiga toxin-producing *Escherichia coli* O26 in Italy and Romania: a new
747 challenge for the European Union. Euro Surveill. 2016;**21**. doi:10.2807/1560-
748 7917.ES.2016.21.49.30420

749 24 Loconsole D, Giordano M, Centrone F, *et al.* Epidemiology of Shiga toxin-producing
750 *Escherichia coli* infections in Southern Italy after implementation of symptom-based surveillance
751 of bloody diarrhea in the pediatric population. Int J Env. Res Public Health. 2020;**17**.
752 doi:10.3390/ijerph17145137

753 25 Ogura Y, Gotoh Y, Itoh T, *et al.* Population structure of *Escherichia coli* O26:H11 with recent
754 and repeated *stx2* acquisition in multiple lineages. Microb Genom. 2017;**3**.
755 doi:10.1099/mgen.0.000141

This is a provisional file, not the final typeset article

756 26 Fratamico PM, Yan X, Caprioli A, *et al.* The complete DNA sequence and analysis of the
757 virulence plasmid and of five additional plasmids carried by Shiga toxin-producing *Escherichia*
758 *coli* O26:H11 strain H30. *Int J Med Microbiol* 2011;**301**:192–203.
759 doi:10.1016/j.ijmm.2010.09.002

760 27 Leomil L, Pestana de Castro AF, Krause G, *et al.* Characterization of two major groups of
761 diarrheagenic *Escherichia coli* O26 strains which are globally spread in human patients and
762 domestic animals of different species. *FEMS Microbiol Lett* 2005;**249**:335–42.
763 doi:10.1016/j.femsle.2005.06.030

764 28 Bielaszewska M, Prager R, Kock R, *et al.* Shiga toxin gene loss and transfer in vitro and in vivo
765 during enterohemorrhagic *Escherichia coli* O26 infection in humans. Appl Env. Microbiol.
766 2007;**73**:3144–50. doi:10.1128/AEM.02937-06

767 29 Bugarel M, Beutin L, Scheutz F, *et al.* Identification of genetic markers for differentiation of
768 Shiga Toxin-producing, enteropathogenic, and avirulent strains of *Escherichia coli* O26. *Appl*
769 *Environ Microbiol* 2011;**77**:2275–81. doi:10.1128/AEM.02832-10

770 30 Delannoy S, Beutin L, Fach P. Discrimination of enterohemorrhagic *Escherichia coli* (EHEC)
771 from non-EHEC strains based on detection of various combinations of type III effector genes. *J*
772 *Clin Microbiol* 2013;**51**:3257–62. doi:10.1128/JCM.01471-13

773 31 Vlisidou I, Marches O, Dziva F, *et al.* Identification and characterization of EspK, a type III
774 secreted effector protein of enterohaemorrhagic *Escherichia coli* O157:H7. FEMS Microbiol
775 Lett. 2006;**263**:32–40. doi:10.1111/j.1574-6968.2006.00410.x

776 32 Steyert SR, Rasko DA, Kaper JB. Functional and phylogenetic analysis of *ureD* in Shiga toxin-
777 producing *Escherichia coli*. *J Bacteriol* 2011;**193**:875–86. doi:10.1128/JB.00922-10

778 33 Delannoy S, Beutin L, Fach P. Towards a molecular definition of enterohemorrhagic *Escherichia*
779 *coli* (EHEC): Detection of genes located on O Island 57 as markers to distinguish EHEC from
780 closely related enteropathogenic *E. coli* strains. *J Clin Microbiol* 2013;**51**:1083–8.
781 doi:10.1128/JCM.02864-12

782 34 Food Standards Scotland. Whole Genome Sequence Typing and Analysis of Non-O157 STEC.
783 Published Online First: 2020.https://www.foodstandards.gov.scot/publications-and-
784 research/publications/whole-genome-sequence-typing-and-analysis-of-non-o157-stec (accessed
785 14 May 2021).

786 35 Public Health Scotland. STEC in Scotland, 2019. Enhanced surveillance and reference laboratory
787 data - STEC in Scotland: enhanced surveillance and reference laboratory data. Published Online
788 First: 2020.https://publichealthscotland.scot/publications/stec-in-scotland-enhanced-surveillance-
789 and-reference-laboratory-data/stec-in-scotland-2019-enhanced-surveillance-and-reference-
790 laboratory-data/ (accessed 12 Dec 2022).

791 36 Pearce MC, Evans J, McKendrick IJ, *et al.* Prevalence and virulence factors of *Escherichia coli*
792 serogroups O26, O103, O111, and O145 shed by cattle in Scotland. Appl Env. Microbiol.
793 2006;**72**:653–9. doi:10.1128/AEM.72.1.653-659.2006

794 37 Pearce MC, Chase-Topping ME, McKendrick IJ, *et al.* Temporal and spatial patterns of bovine
795 *Escherichia coli* O157 prevalence and comparison of temporal changes in the patterns of phage
796 types associated with bovine shedding and human *E. coli* O157 cases in Scotland between 1998-
797 2000 and 2002-2004. BMC Microbiol. 2009;**9**:276. doi:10.1186/1471-2180-9-276

798 38 Henry MK, Tongue SC, Evans J, *et al.* British *Escherichia coli* O157 in Cattle Study (BECS): to
799 determine the prevalence of *E. coli* O157 in herds with cattle destined for the food chain.
800 Epidemiol Infect. 2017;**145**:3168–79. doi:10.1017/S0950268817002151

801 39 Hoyle DV, Keith M, Williamson H, *et al.* Prevalence and epidemiology of Non-O157
802 *Escherichia coli* Serogroups O26, O103, O111, and O145 and Shiga toxin gene carriage in
803 Scottish cattle, 2014–2015. *Appl Environ Microbiol* 2021;**87**:e03142-20.
804 doi:10.1128/AEM.03142-20

805 40 Dallman TJ, Greig DR, Gharbia SE, *et al.* Phylogenetic context of Shiga toxin-producing
806 *Escherichia coli* serotype O26:H11 in England. Microb Genom. 2021;**7**.
807 doi:10.1099/mgen.0.000551

808 41 Rodwell EV, Simpson A, Chan Y-W, *et al.* The epidemiology of Shiga toxin-producing
809 Escherichia coli O26:H11 (clonal complex 29) in England, 2014–2021. *J Infect* 2023;**86**:552–62.
810 doi:10.1016/j.jinf.2023.04.006

811 42 Zhou Z, Alikhan N-F, Mohamed K, *et al.* The EnteroBase user's guide, with case studies on
812 *Salmonella transmissions*, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity.
813 *Genome Res* 2020;**30**:138–52. doi:10.1101/gr.251678.119

814 43 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
815 *Bioinformatics* 2014;**30**:2114–20. doi:10.1093/bioinformatics/btu170

816 44 Chattaway MA, Schaefer U, Tewolde R, *et al.* Identification of Escherichia coli and Shigella
817 Species from Whole-Genome Sequences. *J Clin Microbiol* 2017;**55**:616–23.
818 doi:10.1128/JCM.01790-16

819 45 Langmead B, Trapnell C, Pop M, *et al.* Ultrafast and memory-efficient alignment of short DNA
820 sequences to the human genome. *Genome Biol* 2009;**10**:R25. doi:10.1186/gb-2009-10-3-r25

821 46 Tewolde R, Dallman T, Schaefer U, *et al.* MOST: a modified MLST typing tool based on short
822 read sequencing. *PeerJ* 2016;**4**:e2308. doi:10.7717/peerj.2308

823 47 Ashton PM, Perry N, Ellis R, *et al.* Insight into Shiga toxin genes encoded by Escherichia coli
824 O157 from whole genome sequencing. *PeerJ* 2015;**3**:e739. doi:10.7717/peerj.739

825 48 Silva M, Machado MP, Silva DN, *et al.* chewBBACA: A complete suite for gene-by-gene
826 schema creation and strain identification. *Microb Genomics*;**4**:e000166.
827 doi:10.1099/mgen.0.000166

828 49 Bushnell B. BBMap. sourceforge.net/projects/bbmap/

This is a provisional file, not the final typeset article

829 50 Bankevich A, Nurk S, Antipov D, *et al.* SPAdes: a new genome assembly algorithm and its
830 applications to single-cell sequencing. *J Comput Biol* 2012;**19**:455–77.
831 doi:10.1089/cmb.2012.0021

832 51 Treangen TJ, Ondov BD, Koren S, *et al.* The Harvest suite for rapid core-genome alignment and
833 visualization of thousands of intraspecific microbial genomes. *Genome Biol* 2014;**15**:524.
834 doi:10.1186/s13059-014-0524-x

835 52 Croucher NJ, Page AJ, Connor TR, *et al.* Rapid phylogenetic analysis of large samples of
836 recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015;**43**:e15.
837 doi:10.1093/nar/gku1196

838 53 Minh BQ, Schmidt HA, Chernomor O, *et al.* IQ-TREE 2: New models and efficient methods for
839 phylogenetic inference in the Genomic Era. *Mol Biol Evol* 2020;**37**:1530–4.
840 doi:10.1093/molbev/msaa015

841 54 Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display
842 and annotation. *Nucleic Acids Res* 2021;**49**:W293–6. doi:10.1093/nar/gkab301

843 55 Seemann T. ABRicate. 2022.https://github.com/tseemann/abricate (accessed 13 Dec 2022).

844 56 *Escherichia coli* virulence factors. 2021.https://github.com/phac-nml/ecoli_vf (accessed 13 Dec
845 2022).

846 57 Bharat A, Petkau A, Avery BP, *et al.* Correlation between phenotypic and In Silico detection of
847 Antimicrobial Resistance in *Salmonella enterica* in Canada Using Staramr. *Microorganisms*
848 2022;**10**:292. doi:10.3390/microorganisms10020292

849 58 Florensa AF, Kaas RS, Clausen PTLC, *et al.* ResFinder – an open online resource for
850 identification of antimicrobial resistance genes in next-generation sequencing data and prediction
851 of phenotypes from genotypes. *Microb Genomics* 2022;**8**:000748. doi:10.1099/mgen.0.000748

852 59 Johansson MHK, Bortolaia V, Tansirichaiya S, *et al.* Detection of mobile genetic elements
853 associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool:
854 MobileElementFinder. *J Antimicrob Chemother* 2021;**76**:101–9. doi:10.1093/jac/dkaa390

855 60 Carver T, Harris SR, Berriman M, *et al.* Artemis: an integrated platform for visualization and
856 analysis of high-throughput sequence-based experimental data. *Bioinformatics* 2012;**28**:464–9.
857 doi:10.1093/bioinformatics/btr703

858 61 Camacho C, Coulouris G, Avagyan V, *et al.* BLAST+: architecture and applications. *BMC*
859 *Bioinformatics* 2009;**10**:421. doi:10.1186/1471-2105-10-421

860 62 The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids*
861 *Res* 2021;**49**:D480–9. doi:10.1093/nar/gkaa1100

862 63 Arndt D, Grant JR, Marcu A, *et al.* PHASTER: a better, faster version of the PHAST phage
863 search tool. *Nucleic Acids Res* 2016;**44**:W16–21. doi:10.1093/nar/gkw387

864 64 Seemann T. Snippy. 2022.https://github.com/tseemann/snippy (accessed 13 Dec 2022).

865 65 Disty McMatrixface. 2021.https://github.com/c2-d2/disty (accessed 13 Dec 2022).

866 66 Dallman TJ, Byrne L, Ashton PM, *et al.* Whole-Genome Sequencing for National Surveillance of
867 Shiga Toxin–Producing Escherichia coli O157. *Clin Infect Dis* 2015;**61**:305–12.
868 doi:10.1093/cid/civ318

869 67 Long J, Geng J, Xu Y, *et al.* Large-scale phylogenetic analysis reveals a new genetic clade
870 among *Escherichia coli* O26 strains. *Microbiol Spectr* 2022;**10**:e02525-21.
871 doi:10.1128/spectrum.02525-21

872 68 Karnisova L, Marejkova M, Hrbackova H, *et al.* Attack of the clones: whole genome-based
873 characterization of two closely related enterohemorrhagic *Escherichia coli* O26 epidemic
874 lineages. BMC Genomics. 2018;**19**:647. doi:10.1186/s12864-018-5045-7

875 69 Ishijima N, Lee KI, Kuwahara T, *et al.* Identification of a new virulent clade in
876 enterohemorrhagic *Escherichia coli* O26:H11/H- Sequence Type 29. Sci Rep. 2017;**7**:43136.
877 doi:10.1038/srep43136

878 70 Gonzalez-Escalona N, Toro M, Rump LV, *et al.* Virulence Gene Profiles and Clonal
879 Relationships of *Escherichia coli* O26:H11 Isolates from feedlot cattle as determined by Whole-
880 Genome Sequencing. Appl Env. Microbiol. 2016;**82**:3900–12. doi:10.1128/AEM.00498-16

881 71 Browne AS, Midwinter AC, Withers H, *et al.* Molecular epidemiology of Shiga Toxin-Producing
882 *Escherichia coli* (STEC) on New Zealand dairy farms: Application of a culture-independent
883 assay and Whole-Genome Sequencing. Appl Env. Microbiol. 2018;**84**. doi:10.1128/aem.00481-
884 18

885 72 Browne AS, Biggs PJ, Wilkinson DA, *et al.* Use of genomics to investigate historical importation
886 of Shiga Toxin-Producing *Escherichia coli* Serogroup O26 and nontoxigenic variants into New
887 Zealand. Emerg Infect Dis. 2019;**25**:489–500. doi:10.3201/eid2503.180899

888 73 Pukatzki S, McAuley SB, Miyata ST. The type VI secretion system: translocation of effectors
889 and effector-domains. *Curr Opin Microbiol* 2009;**12**:11–7. doi:10.1016/j.mib.2008.11.010

890 74 Michelacci V, Montalbano Di Filippo M, Gigliucci F, *et al.* Population analysis of O26 Shiga
891 toxin-producing *Escherichia coli* causing hemolytic uremic syndrome in Italy, 1989-2020,
892 through whole genome sequencing. *Front Cell Infect Microbiol* 2022;**12**:842508.
893 doi:10.3389/fcimb.2022.842508

894 75 Senthakumaran T, Brandal LT, Lindstedt BA, *et al.* Implications of stx loss for clinical
895 diagnostics of Shiga toxin-producing *Escherichia coli*. Eur J Clin Microbiol Infect Dis.
896 2018;**37**:2361–70. doi:10.1007/s10096-018-3384-6

897 76 Tauschek M, Gorrell RJ, Strugnell RA, *et al.* Identification of a protein secretory pathway for the
898 secretion of heat-labile enterotoxin by an enterotoxigenic strain of *Escherichia coli*. *Proc Natl*
899 *Acad Sci* 2002;**99**:7066–71. doi:10.1073/pnas.092152899

900 77 Yang J, Baldi DL, Tauschek M, *et al.* Transcriptional Regulation of the *yghJ-pppA-yghG-*
901 *gspCDEFGHIJKLM* cluster, encoding the Type II Secretion Pathway in enterotoxigenic
902 *Escherichia coli*. *J Bacteriol* 2007;**189**:142–50. doi:10.1128/JB.01115-06

This is a provisional file, not the final typeset article

903 78 Strozen TG, Li G, Howard SP. YghG (GspSβ) is a novel pilot protein required for localization of
904 the GspSβ Type II Secretion System secretin of enterotoxigenic *Escherichia coli*. *Infect Immun*
905 2012;**80**:2608–22. doi:10.1128/IAI.06394-11

906 79 Lu X, Fu E, Xie Y, *et al.* Electron acceptors induce secretion of enterotoxigenic *Escherichia coli*
907 heat-labile enterotoxin under anaerobic conditions through promotion of GspD assembly. *Infect*
908 *Immun* 2016;**84**:2748–57. doi:10.1128/IAI.00358-16

909 80 Wang H, Sanz Garcia R, Cox E, *et al.* Porcine enterotoxigenic *Escherichia coli* strains differ in
910 their capacity to secrete enterotoxins through varying YghG levels. *Appl Environ Microbiol*
911 2020;**86**:e00523-20. doi:10.1128/AEM.00523-20

912 81 Ristow LC, Welch RA. Hemolysin of uropathogenic *Escherichia coli*: A cloak or a dagger?
913 *Biochim Biophys Acta BBA - Biomembr* 2016;**1858**:538–45. doi:10.1016/j.bbamem.2015.08.015

914 82 Bernard P, Couturier M. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-
915 topoisomerase II complexes. *J Mol Biol* 1992;**226**:735–45. doi:10.1016/0022-2836(92)90629-X

916 83 Scaletsky ICA, Michalski J, Torres AG, *et al.* Identification and characterization of the locus for
917 diffuse adherence, which encodes a novel afimbrial adhesin found in atypical enteropathogenic
918 *Escherichia coli*. *Infect Immun* 2005;**73**:4753–65. doi:10.1128/IAI.73.8.4753-4765.2005

919 84 Shaikh N, Tarr PI. Escherichia coli O157:H7 Shiga Toxin-Encoding Bacteriophages:
920 Integrations, Excisions, Truncations, and Evolutionary Implications. *J Bacteriol* 2003;**185**:3596–
921 605. doi:10.1128/jb.185.12.3596-3605.2003

922 85 Rodríguez-Rubio L, Haarmann N, Schwidder M, *et al.* Bacteriophages of Shiga Toxin-Producing
923 Escherichia coli and Their Contribution to Pathogenicity. *Pathogens* 2021;**10**:404.
924 doi:10.3390/pathogens10040404

925 86 Hall RJ, Whelan FJ, Cummins EA, *et al.* Gene-gene relationships in an *Escherichia coli*
926 accessory genome are linked to function and mobility. *Microb Genomics* 2021;**7**:000650.
927 doi:10.1099/mgen.0.000650

928 87 Gonzalez-Escalona N, Allard MA, Brown EW, *et al.* Nanopore sequencing for fast determination
929 of plasmids, phages, virulence markers, and antimicrobial resistance genes in Shiga toxin-
930 producing *Escherichia coli*. PLoS One. 2019;**14**:e0220494. doi:10.1371/journal.pone.0220494

931 88 Critically important antimicrobials for human medicine, 6th revision. Geneva: : World Health
932 Organization 2019.

933 89 Day M, Doumith M, Jenkins C, *et al.* Antimicrobial resistance in Shiga toxin-producing
934 *Escherichia coli* serogroups O157 and O26 isolated from human cases of diarrhoeal disease in
935 England, 2015. *J Antimicrob Chemother* 2017;**72**:145–52. doi:10.1093/jac/dkw371

936 90 Humphry RW, Henry MK, Reeves A, *et al.* Estimating antimicrobial usage based on sales to beef
937 and dairy farms from UK veterinary practices. *Vet Rec* 2021;**189**:e28. doi:10.1002/vetr.28

938 91 RUMA. Antibiotic Use In Cattle: Dairy Trend Data and Usage by Beef Farm Type 2015-2019.
939 RUMA 2021. https://ruma.org.uk/wp-content/uploads/2022/07/Antibiotic-Use-in-936 Cattle-
940 Report.pdf (accessed 11 Jul 2023).

941 92 Tarr PI, Freedman SB. Why antibiotics should not be used to treat Shiga toxin-producing
942 Escherichia coli infections. *Curr Opin Gastroenterol* 2022;**38**:30.
943 doi:10.1097/MOG.0000000000000798

944

945

946 **SUPPLEMENTARY MATERIAL**

947 Supplementary Material is provided as a .zip file.

This is a provisional file, not the final typeset article