

1 **The long polar fimbriae (*lpf*) operon and its flanking regions in bovine**  
2 ***Escherichia coli* O157:H43 and STEC O136:H12 strains**

3

4 Domonkos Sváb<sup>1</sup>, Lucia Galli<sup>2</sup>, Balázs Horváth<sup>3,4</sup>, Gergely Maróti<sup>3,4</sup>, Ulrich Dobrindt<sup>5</sup>,  
5 Alfredo G. Torres<sup>6</sup>, Marta Rivas<sup>2</sup>, István Tóth<sup>1\*</sup>

6

7 <sup>1</sup>Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian  
8 Academy of Sciences, Budapest, Hungary

9 <sup>2</sup>Servicio Fisiopatogenia, Departamento de Bacteriología, Instituto Nacional de Enfermedades  
10 Infecciosas-ANLIS “Dr. Carlos G. Malbrán”, Buenos Aires, Argentina

11 <sup>3</sup>Bay Zoltán Nonprofit Research Ltd., Szeged, Hungary

12 <sup>4</sup>Biological Research Centre, Hungarian Academy of Sciences, Szeged, Hungary

13 <sup>5</sup>Institute for Hygiene, University of Münster, Münster, Germany

14 <sup>6</sup>Department of Microbiology and Immunology, University of Texas Medical Branch,  
15 Galveston, Texas 77555-1070

16

17 \*Corresponding author. Institute for Veterinary Medical Research, Centre for Agricultural  
18 Research, Hungarian Academy of Sciences, H-1143, Hungária krt. 21., Budapest, Hungary

19 E-mail: [tothi@vmri.hu](mailto:tothi@vmri.hu)

20

21 **Running title:** Lpf in atypical *E. coli* O157 and STEC O136 strains

22

23 **Keywords:** *Escherichia coli*, O157:H43, atypical *E. coli*, pathogenic *E. coli*, long polar  
24 fimbriae, O136:H12

25 **ABSTRACT**

26

27 Long polar fimbriae (Lpf) are intestinal adhesins and important virulence factors of  
28 pathogenic *Escherichia coli* strains. We cloned and sequenced the *lpf2-1* operon (*lpf2ABCD*)  
29 and its flanking regions of an intimin- and Shiga toxin-negative *E. coli* O157:H43 strain from  
30 bovine origin, and also sequenced the *lpf2-1* operon of 6 additional atypical O157 bovine  
31 *Escherichia coli* strains of various serotypes Nucleotide sequence comparison of these *lpf*  
32 operons showed sequence conservation as they contain only four polymorphic nucleotide  
33 positions. Investigation of these O157 strains as well as 13 *Escherichia coli* Reference  
34 Collection (ECOR) strains carrying the *lpf2-1* allele revealed high degree of sequence  
35 conservation in the *lpf2* flanking regions. The *lpf2-1* allele is also present in a bovine Shiga  
36 toxin-producing *E. coli* STEC O136:H12 strain and *in vitro* adherence assays revealed that the  
37 absence of *lpf2-1* in this strain did not affect its host cell-binding properties. Our data indicate  
38 that *lpf2* loci is highly conserved in *E. coli* isolates, but its role in adherence might be masked  
39 by other uncharacterized adhesins.

40

41

## 42 INTRODUCTION

43

44 *Escherichia coli* is an important member of the commensal intestinal microflora in mammals,  
45 but there are a large number of isolates, which have acquired a variety of virulence factors and  
46 are capable of causing serious diseases in humans and animals, including those classified as  
47 enterohemorrhagic *E. coli* (EHEC; Kaper et al., 2004). The frequent emergence of new  
48 isolates with new combinations of virulence genes is exemplified by the appearance of the  
49 strain responsible for the recent outbreak of hemolytic uremic syndrome (HUS) in Germany  
50 (Mellmann et al., 2011), and underlines the importance of the various possible lateral gene  
51 transfer mechanisms in the spread of virulence genes. Typical EHEC O157:H7/NM strains  
52 carry *stx* genes encoding Shiga-toxin and also harbour a pathogenicity island (PAI) known as  
53 the locus of enterocyte effacement (LEE), encoding the intimin adhesin, among other  
54 virulence factors (Kaper et al., 2004).

55 In addition to several extensively studied virulence factors carried by pathogenic *E.*  
56 *coli*, there are several additional factors such as the long polar fimbriae (Lpf), which represent  
57 a relatively recently described adhesin and virulence determinant in EHEC (Doughty et al.,  
58 2002). The exact mechanism by which Lpf contributes to the virulence of each pathogenic *E.*  
59 *coli* strain is currently under investigation, but there is well-documented evidence that Lpf  
60 promote adhesion of EHEC strains to the intestinal epithelium (Jordan et al., 2004; Fitzhenry  
61 et al., 2006; Torres et al., 2008) as well as showing that Lpf interacts with extracellular matrix  
62 proteins (Farfan et al., 2011).

63 Initially, two genetic variants of Lpf (Lpf1 and Lpf2) were identified in *E. coli*, and  
64 with the availability of additional sequence data, more variants have been discovered, some of  
65 which show a degree of association with certain serotypes and/or pathogroups (Torres et al.,  
66 2009). All known *lpf1* and *lpf2* operons are encoded on genomic islands, termed O islands,

67 integrated in specific chromosomal locations (Fitzhenry et al., 2006; Mellmann et al., 2011).  
68 The Lpf variant encoded by the operon first named *lpfA<sub>O113</sub>* (Ideses et al., 2005) and later  
69 termed as allele 1 of *lpf2* (*lpf2-1*) (Torres et al., 2009) is the most prevalent genetic variant of  
70 Lpf according to our present knowledge, as it has been detected in several strains from  
71 various serotypes (Toma C, 2004; Toma et al., 2006; Torres et al., 2009; Galli et al., 2010;  
72 Monaghan et al., 2011). In comparison, the integration site of the *lpf2* operon is found  
73 between the genes coding for the L-glutamine:D-fructose-6-phosphate aminotransferase  
74 (*glmS*) and that of phosphate-binding periplasmic protein (*pstS*). Recently, seven *E. coli*  
75 strains of the O157 serogroup (including strain T22), isolated from healthy cattle and without  
76 key EHEC virulence factors were found to harbour the Lpf2 variant (Sváb & Tóth, 2012).  
77 Three isolates were from the serotype O157:H43 (Tóth et al., 2009), and some of these  
78 members have been the focus of a recent study dealing with the evolution of the O157  
79 serogroup (Iguchi et al., 2011).

80         In the current study, we report the sequence of the *lpf2* operon and its flanking regions  
81 in strain T22, a non-sorbitol-fermenting (NSF) O157:H43 with an atypical pathotype (*stx*-,  
82 *eae*), monitor the presence of the operon and its flanking regions in a collection of *E. coli*  
83 strains of various serotypes carrying the *lpf2-1* allele, and investigate the possible function of  
84 Lpf2 in adherence *in vitro* of a STEC O136:H12, that does not possess intimin or other known  
85 adhesins found in bovine STEC strains.

86 **MATERIAL AND METHODS**

87

88 **Bacterial strains.** *E. coli* strains used in this study are listed in Table 1. The ECOR strains  
89 were provided by Mónika Kerényi (Department of Medical Microbiology and Immunology,  
90 Medical School, University of Pécs, Hungary). Strains were grown in lysogeny broth (LB), as  
91 well as on LB and bromothymol-blue agar plates. For isolation of genomic and cosmid DNA,  
92 strains were grown in tryptic soy broth (TSB).

93

94

95 **Cosmid clone library construction.** Genomic DNA was isolated from strain T22 with the  
96 phenol-chlorophorm method (Sambrook et al., 1982) after growing overnight in TSB. The  
97 preparation of the cosmid clone library was performed with pWEB-TNC Cosmid Cloning Kit  
98 (Epicentre, Madison, WI, USA) according to the manufacturer's instructions, with the  
99 modification that instead of mechanical shearing, genomic DNA was subjected to a partial  
100 digestion with restriction endonuclease *Mbo*I (Fermentas, Vilnius, Lithuania). Altogether,  
101 1000 transformant colonies were kept as cosmid library.

102

103 **PCR screening for the presence of *lpf2* flanking regions.** The cosmid library was screened  
104 by PCR for the presence of *lpf2*. The primers and annealing temperatures used in the reactions  
105 are listed in Table 2. The strains listed in Table 1 were screened for the presence of flanking  
106 regions.

107

108 **Reverse Transcriptase PCR for *lpfA* of *E. coli* T22.** RNA was isolated from cells of a 48 h  
109 culture of *E. coli* strain T22 with RNEasy mini kit (Qiagen, Hilden, Germany) according to  
110 the manufacturer's instructions, with the modification that cells were collected with

111 centrifugation at 13,000 xg for 1 minute. After discarding the supernatant, the bacterial pellet  
112 was processed according to the manufacturers' instructions. RNA samples were treated with  
113 Sigma DNase I Amplification Grade (Sigma-Aldrich, St. Louis, MO, USA). The DNase  
114 treated samples were used as template for reverse transcription with Fermentas Maxima  
115 Reverse Transcriptase (Fermentas, Vilnius, Lithuania) according to the manufacturer's  
116 protocol. The product of this reaction was used as template in a regular PCR with the primers  
117 defined in Table 2.

118

119 **Sequencing.** A cosmid carrying the whole *lpf* operon was identified. DNA was isolated with  
120 the Sigma GenElute BAC DNA kit (Sigma-Aldrich, St. Louis, MO, USA), and was  
121 sequenced at Baygen Institute (Szeged, Hungary) using the combination of Life Tech's  
122 SOLiD 4, IonTorrent sequencing and the dideoxynucleotide methods. The products of the  
123 RT-PCR were also sequenced with the dideoxynucleotide method. Nucleotide sequence  
124 analysis and searches for open reading frames (ORFs) and homologous DNA sequences in the  
125 EMBL and GenBank database libraries were performed with the tools available from the  
126 National Center for Biotechnology Information ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), with Vector NTI  
127 and the CLC Bio DNA Workbench.

128

129 **Construction of an *lpfA2-1* deletion mutant of STEC O136:H12.** To generate an *lpfA2-1*  
130 deletion mutant of strain 187/06 (22), the *lpfA2-1* gene was replaced by a gene encoding  
131 kanamycin resistance using the lambda red recombinase system (Datsenko & Wanner, 2000).  
132 The long oligonucleotide primers used for introducing the mutation were those described by  
133 (Doughty et al., 2002). Each primer included 20 bp of sequence homologous to the  
134 kanamycin resistance gene, and 40 bp of sequence homologous to regions flanking the *lpfA2-*  
135 *I* gene. The kanamycin resistance gene was amplified from pKD4 by PCR. The purified PCR

136 product (1 µg DNA) was electroporated into 187/06 (22) strain which had previously been  
137 transformed with the lambda red recombinase expression vector, pKM201. Following  
138 electroporation, transformants of 187/06 (22) were recovered at 30°C for 2 h in LB broth and  
139 plated onto LB agar with kanamycin for overnight growth at 37°C to induce the loss of  
140 pKM201. Kanamycin-resistant colonies were then confirmed by PCR for replacement of  
141 *lpfA2-1*. The *lpfA2-1* deletion mutant of 187/06 (22) was complemented in trans by  
142 introduction of the entire *lpf2* operon on pWSK:lpf (kindly provided by Dr. E. Hartland).

143

144 **Bacterial adhesion.** The potential adherence capacity of strain T22 was investigated on  
145 primary bovine kidney and testis cells, which were kindly provided by Emília Szállás  
146 (Veterinary Diagnostic Directorate, National Food Chain Safety Office, Budapest, Hungary).  
147 The cells were grown to semi-confluency at 37°C in 5% CO<sub>2</sub> in 24 well plates in Roswell  
148 Park Memorial Institute 1640 (RPMI 1640) medium without supplements. Prior to use, cells  
149 were washed once with PBS. Strain T22 was grown for 48 hours with shaking at 200 rpm,  
150 and the cell monolayers were incubated for 5 hours with ca. 10<sup>10</sup> bacteria per well. The  
151 infected monolayers were washed two times with PBS, fixed with methanol and stained with  
152 Giemsa reagent.

153 The ability of *E. coli* O136:H12 *lpfA2-1*<sup>+</sup> strain and its deletion mutant to adhere to  
154 Hep-2, Caco-2 and T84 cell lines was assessed as previously described (Doughty et al., 2002),  
155 with minor modifications. The cells were grown to semi-confluency at 37°C in 5% CO<sub>2</sub> in 24  
156 well plates (Falcon<sup>TM</sup> BD) in Dulbecco's minimal essential medium (DMEM), DMEM/F12  
157 (Gibco, Carlsbad, CA, USA) or MEM, depending of the cell line, with 10% or 20% (vol/vol)  
158 heat-inactivated fetal bovine serum, 2 mM L-glutamine, and 1% (vol/vol) of a mixture of  
159 antibiotics/antimycotics (Gibco). Before use, the cells were washed twice with phosphate-  
160 buffered saline (PBS, Gibco) and replenished with the corresponding medium with no

161 supplements, as it was observed previously that mannose could inhibit Lpf mediated  
162 adherence (Farfan et al., 2011). The strains were grown static in LB broth overnight at 37°C,  
163 tissue culture cells were incubated with ca.  $10^7$  bacteria per well for 3 h at 37°C and 5% CO<sub>2</sub>.  
164 To quantify adherence, the infected monolayers were washed two times with PBS, and the  
165 adherent bacteria were recovered with 200 µl of 0.1% Triton X-100 in PBS and plated on LB  
166 agar plates containing the proper antibiotic. Data were expressed as the percentage of the  
167 bacterial inoculum recovered from triplicate wells and are the mean of at least two separate  
168 experiments. Statistical difference was expressed as the *P* value determined by a *t* test  
169 analysis. The *in vitro* competition assays were performed as described above except that cells  
170 were inoculated with  $5 \times 10^6$  cells each of mutant and wild type bacteria (total number of  
171 bacteria/well  $10^7$  cells) and competition index (CI) was calculated.

172

173 **Nucleotide sequence accession number.** The sequence of the *E. coli* T22 *lpf2* operon and its  
174 neighbouring regions has been deposited in the GenBank database under accession number  
175 AHZD01000104. The sequences of the *lpf2* operons of *E. coli* O157 strains B47, B54, T16,  
176 T34, T49 and T50 (see Table 1) were deposited under accession numbers KC207119,  
177 KC207120, KC207121, KC207122, KC207123 and KC207124, respectively.

178 **RESULTS AND DISCUSSION**

179

180 **Sequence characteristics of the *lpf2* operon in the atypical bovine O157 strains.** We  
181 cloned and sequenced a 15.3 kb region from the genome of *E. coli* O157:H43 strain T22,  
182 including the *lpf2* operon. The sequence is deposited in GenBank under the accession number  
183 AHZD01000104. The schematic representation of the sequenced region is shown in Figure 1.  
184 According to our knowledge, this is the first time that the allelic variant *lpf2-1* operon and its  
185 flanking regions were sequenced in a non-sorbitol fermenting (NSF) strain of the serotype  
186 O157:H43. The *lpf2-1* operon itself has been detected earlier by PCR in other *E. coli* strains  
187 (Torres et al., 2009; Farfan et al., 2011).

188 In the *lpf2-1* operons of T22 and six other atypical O157 strains, there are only 4  
189 positions that show polymorphism. One of them is a synonymous point mutation in the gene  
190 *lpfC* of strain T49, the others produce amino acid switches in the respective genes. The *lpf2A*  
191 gene is uniformly conserved in the investigated strains, and this is also true for the majority of  
192 the strains with whole genomes available in GenBank, an exception is the *lpf2A* gene of strain  
193 55989 (Table 3.), which contains an isoleucine instead of a leucine in position 116. The *lpf2B*  
194 has an alanine instead of serine in position 99 in all the atypical O157 strains as opposed to  
195 the strains from GenBank listed in Table 3.

196 The *lpf2C* gene proved to be uniform in all the sequenced strains with the exception of  
197 T49, which has a cysteine in position 809 instead of tryptophane – the second polymorphism  
198 within the *lpf* operons of the atypical O157 strains. The only strain with an amino acid switch  
199 in *lpf2C* available in GenBank is E24377A, which has a leucine instead of proline in position  
200 122. In the case of *lpf2D* gene, strain T22 has serine instead of alanine in position 341, this  
201 switch is shared with strains SE11, 55989 and 11368, and is the third polymorphism within  
202 the sequenced atypical strains. The fourth polymorphism can be observed in strains T49 and

203 T50, which encode a leucine instead of a methionine in position 313. The nucleotide sequence  
204 comparison of the *lpf* genes of the atypical O157 bovine strains is shown in Figure S1  
205 (Supporting Material).

206 The existence of these polymorphisms indicate a similar level of variation in the  
207 otherwise conserved *lpf2* sequences. However, currently there is no data available on the  
208 potential or actual effect of these differences on the expression and/or function of Lpf2.  
209 The GC content of the sequenced *lpf* operons was 44%, while that of the flanking regions in  
210 T22 was 52%, close to the average GC content in the *E. coli* genome (McLean et al., 1998).  
211 This fact, together with the generally conserved sequence of *lpf2* in strains of various sero-  
212 and pathotypes (Table 3) suggests that these *lpf* variants might be located in genomic islands.

213

214 **Dissemination and characteristics of the flanking regions of *lpf2*.** The results of the PCR  
215 scanning of previously characterized strains carrying allele *lpf2-1* are listed in Table 1. The  
216 fact that the *lpf2-1* operon is flanked by the same set of genes in the majority of the strains  
217 sequenced so far (Table 3) is further demonstrated by the results of our PCR scanning. This  
218 uniformity indicates that the site between the *pstS* and *glmS* genes served as an integration  
219 hotspot at some point during the evolution of these strains. The genetic analysis of this region  
220 was performed in an earlier study, in which the authors designed primers specific for the  
221 flanking regions, and investigated whether the site between *pstS* and *glmS* is intact or  
222 interrupted by the *lpf2* operon (Toma et al., 2006). It must be noted however, that in the case  
223 of prototypic enteroaggregative strain *E. coli* (EAEC) strain 042 (O44:H18), a *Tn21*  
224 transposon sequence is inserted between *lpfA* and *glmS* genes. This transposon element  
225 contains among other features transposases and genes encoding antibiotic resistance  
226 (Chaudhuri et al., 2010). Interestingly, four out of five strains which have the closest  
227 homologues to the *lpf2-1* operon of strain T22, are commensal isolates (Table 3).

228

229 **Expression of Lpf2.** The RT-PCR specific for the *lpfA* genes from T22 yielded positive  
230 results, confirming transcription of Lpf2 in 48-hour cultures. In an earlier study with EHEC  
231 strain EDL933, one of the authors of the current manuscript found that the H-NS protein has a  
232 silencer role, while the regulatory protein Ler acts as an anti-silencer during the expression of  
233 Lpf1 (Torres et al., 2008). Our findings, as well as the lack of LEE (which includes the *ler*  
234 gene) in strain T22 suggests a different regulatory mechanism controlling Lpf2 expression  
235 relative to Lpf1.

236

237 **Contribution of LpfA2-1 to adherence of STEC O136:H12 *in vitro*.** The wide distribution  
238 of this particular allele of Lpf in pathogenic *E. coli* strains, especially in LEE-negative strains  
239 (Doughty et al., 2002) underlines its potential role as an important adhesin. There is both *in*  
240 *vitro* (Doughty et al., 2002; Newton et al., 2004; Torres et al., 2008; Farfan et al., 2011) and *in*  
241 *vivo* (Jordan et al., 2004) experimental evidence that Lpf enhances the adherence of intimin-  
242 positive and negative strains. Construction of an *lpf2* mutant in strain T22 resulted more  
243 challenging than expected; therefore, we created an *lpf2-1* mutant in STEC O136:H12 strain  
244 187/06 (22), a bovine isolate that possesses the same *lpf2-1* allele as strain T22 (data not  
245 shown) and flanked with the same genes (Table 3). The role of Lpf2 in adherence of strain  
246 187/06 (22) was evaluated using different tissue cultured cells lines; however, no clear  
247 differences could be observed in the quantitative adherence assays between O136:H12 and its  
248 corresponding *lpfA2-1* mutant. The strain 187/06 (22) did not show a significant reduction in  
249 the adherence neither to Hep-2 ( $P=0.10$ ), nor to Caco-2 ( $P=0.42$ ) cell lines when was  
250 compared to its deletion mutant, but exhibited a significant reduction in adherence to T84 cell  
251 line ( $P<0.0002$ ) (Figure 2). However, when competition adhesion assays were performed  
252 using the wild type and its corresponding *lpfA2-1* deletion mutant, the mean CI ( $3.27 \pm 1.35$ )

253 was significantly greater than 1 ( $P=0.043$ ). This finding, where the mutant strain is adhering  
254 more than the wild type suggested that the strain lacking Lpf might produce additional  
255 adhesion factors that provided a subtle advantage in the *in vitro* adherence assay.

256 In the case of strain T22, no specific adhesion could be observed in bovine testis and  
257 kidney cell cultures (data not shown), which is in harmony with the finding that strains  
258 possessing Lpf2 are not defective in adhesion but further analysis is needed to define the role  
259 of this fimbriae or other adhesion factors in these subset of strains.

260 In summary, we cloned and sequenced for the first time the long polar fimbriae-  
261 encoding operon and its flanking regions in an atypical, NSF O157:H43 *E. coli* strain. The  
262 Lpf operon itself is nearly identical to those of several other pathogenic and non-pathogenic  
263 strains from various serotypes and pathogroups, and represents the genetic variant termed  
264 allele 1 of *lpf2*. The integration site of the operon also shows high similarity to that on the  
265 aforementioned strains. Characteristic flanking regions were also found in other O157 non H7  
266 strains carrying the same operon. Given the experimental evidence for the role of Lpf in  
267 adherence of *E. coli* O157:H7 and other pathogenic *E. coli* strains, it is plausible to propose  
268 that the Lpf2-1 fimbriae are important factors mediating adherence of our bovine *E. coli*  
269 strains. However, we did not observe any difference in the adherence profile of STEC 187/06  
270 (22) or STEC O136:H12 strains. Based on recent published report, we speculate that in  
271 absence of an adhesin such as Lpf, the strains synthesized alternate surface structure as a  
272 compensatory mechanism for colonization (Lloyd et al., 2012). Further, the complete  
273 regulatory mechanisms of *lpf2* still need to be fully elucidated. Finally, the highly conserved  
274 sequence and integration site of the *lpf2* operon suggests that *lpf2* loci belongs to a conserved  
275 genomic island, and an interesting future task will be to elucidate the mechanism of genetic  
276 acquisition of this operon in pathogenic and commensal *E. coli* strains.

277

278 **ACKNOWLEDGEMENTS**

279 The support of the Hungarian Research Fund (OTKA, grant number K 81252) is  
280 acknowledged. This work was also partially supported by NIH/NIAID AI079154 to AGT. LG  
281 was supported by a fellowship from Consejo Nacional de Investigaciones Científicas y  
282 Técnicas (CONICET). The contents of this work are solely the responsibility of the authors  
283 and do not necessarily represent the official views of the NIAID or NIH.

284

285 **REFERENCES**

286 Appleyard RK (1954) Segregation of New Lysogenic Types during Growth of a Doubly  
287 Lysogenic Strain Derived from *Escherichia coli* K12. *Genetics* **39**:440-452.

288

289 Archer CT, Kim JF, Jeong H, Park JH, Vickers CE, Lee SY, Nielsen LK (2011) The genome  
290 sequence of *E. coli* W (ATCC 9637): comparative genome analysis and an improved genome-  
291 scale reconstruction of *E. coli*. *BMC Genomics* **12**:9.

292

293 Chaudhuri RR, Sebaihia M, Hobman JL, Webber MA, Leyton DL, Goldberg MD,  
294 Cunningham AF, Scott-Tucker A, Ferguson PR, Thomas CM, Frankel G, Tang CM, Dudley  
295 EG, Roberts IS, Rasko DA, Pallen MJ, Parkhill J, Nataro JP, Thomson NR, Henderson IR  
296 (2010) Complete genome sequence and comparative metabolic profiling of the prototypical  
297 enteroaggregative *Escherichia coli* strain 042. *PLoS One* **5**:e8801.

298

299 Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia*  
300 *coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**:6640-6645.

301

302 Doughty S, Sloan J, Bennett-Wood V, Robertson M, Robins-Browne RM, Hartland EL  
303 (2002) Identification of a novel fimbrial gene cluster related to long polar fimbriae in locus of  
304 enterocyte effacement-negative strains of enterohemorrhagic *Escherichia coli*. *Infect Immun*  
305 **70**:6761-6769.

306

307 Farfan MJ, Cantero L, Vidal R, Botkin DJ, Torres AG (2011) Long Polar Fimbriae of  
308 Enterohemorrhagic *Escherichia coli* O157:H7 Bind to Extracellular Matrix Proteins. *Infect*  
309 *Immun* **79**:3744-3750.  
310  
311 Fitzhenry R, Dahan S, Torres AG, Chong Y, Heuschkel R, Murch SH, Thomson M, Kaper  
312 JB, Frankel G, Phillips AD (2006) Long polar fimbriae and tissue tropism in *Escherichia coli*  
313 O157:H7. *Microbes Infect* **8**:1741-1749.  
314  
315 Galli L, Torres AG, Rivas M (2010) Identification of the long polar fimbriae gene variants in  
316 the locus of enterocyte effacement-negative Shiga toxin-producing *Escherichia coli* strains  
317 isolated from humans and cattle in Argentina. *FEMS Microbiol Lett* **308**:123-129.  
318  
319 Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han CG, Ohtsubo E,  
320 Nakayama K, Murata T, Tanaka M, Tobe T, Iida T, Takami H, Honda T, Sasakawa C,  
321 Ogasawara N, Yasunaga T, Kuhara S, Shiba T, Hattori M, Shinagawa H (2001) Complete  
322 genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison  
323 with a laboratory strain K-12. *DNA Res* **8**:11-22.  
324  
325 Ideses D, Biran D, Gophna U, Levy-Nissenbaum O, Ron EZ (2005) The *lpf* operon of  
326 invasive *Escherichia coli*. *Int J Med Microbiol* **295**:227-236.  
327  
328 Iguchi A, Shirai H, Seto K, Ooka T, Ogura Y, Hayashi T, Osawa K, Osawa R (2011) Wide  
329 Distribution of O157-Antigen Biosynthesis Gene Clusters in *Escherichia coli*. *PLoS One*  
330 **6**:e23250.

331 Jordan DM, Cornick N, Torres AG, Dean-Nystrom EA, Kaper JB, Moon HW (2004) Long  
332 polar fimbriae contribute to colonization by *Escherichia coli* O157:H7 *in vivo*. *Infect Immun*  
333 **72**:6168-6171.

334

335 Kaper JB, Nataro JP, Mobley HL (2004) Pathogenic *Escherichia coli*. *Nat Rev Microbiol*  
336 **2**:123-140.

337

338 Lloyd S, Ritchie J, Torres AG (2012) Fimbriation and curliation in *Escherichia coli* O157:H7:  
339 A paradigm of intestinal and environmental colonization. *Gut Microbes* **3**: 272-6.

340

341 McLean MJ, Wolfe KH, Devine KM (1998) Base composition skews, replication orientation,  
342 and gene orientation in 12 prokaryote genomes. *J Mol Evol* **47**:691-696.

343

344 Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K,  
345 Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska  
346 M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Karch H (2011) Prospective  
347 genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4  
348 outbreak by rapid next generation sequencing technology. *PLoS One* **6**:e22751.

349

350 Monaghan A, Byrne B, Fanning S, Sweeney T, McDowell D, Bolton DJ (2011) Serotypes  
351 and Virulence Profiles of non-O157 Shiga-Toxin Producing *Escherichia coli* (STEC) from  
352 Bovine farms. *Appl Environ Microbiol* **77**:8662-8668.

353

354 Newton HJ, Sloan J, Bennett-Wood V, Adams LM, Robins-Browne RM, Hartland EL (2004)  
355 Contribution of long polar fimbriae to the virulence of rabbit-specific enteropathogenic  
356 *Escherichia coli*. *Infect Immun* **72**:1230-1239.  
357  
358 Ochman H, Selander RK (1984) Standard reference strains of *Escherichia coli* from natural  
359 populations. *J Bacteriol* **157**:690-693.  
360  
361 Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H,  
362 Nakayama K, Kurokawa K, Tobe T, Hattori M, Hayashi T (2009) Comparative genomics  
363 reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic  
364 *Escherichia coli*. *Proc Natl Acad Sci U S A* **106**:17939-17944.  
365  
366 Oshima K, Toh H, Ogura Y, Sasamoto H, Morita H, Park S, Ooka T, Iyoda S, Taylor TD,  
367 Hayashi T, Itoh K, Hattori M (2008) Complete genome sequence and comparative analysis of  
368 the wild-type commensal *Escherichia coli* strain SE11 isolated from a healthy adult. *DNA Res*  
369 **15**:375-386.  
370  
371 Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, Gajer P, Crabtree J,  
372 Sebahia M, Thomson NR, Chaudhuri R, Henderson IR, Sperandio V, Ravel J (2008) The  
373 pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal  
374 and pathogenic isolates. *J Bacteriol* **190**:6881-6893.  
375  
376 Sambrook J, Fritsch EF, Maniatis T (1982) *Molecular Cloning: A Laboratory Manual*. Cold  
377 Spring Harbor Laboratory Press, New York.  
378

379 Sváb D, Tóth I (2012) Allelic types of long polar fimbriae in human and bovine *Escherichia*  
380 *coli* O157 strains. *Acta Vet. Hung.* **60**:1-15.

381

382 Toma C, Higa N, Iyoda S, Rivas M, Iwanaga M (2006) The long polar fimbriae genes  
383 identified in Shiga toxin-producing *Escherichia coli* are present in other diarrheagenic *E. coli*  
384 and in the standard *E. coli* collection of reference (ECOR) strains. *Res Microbiol* **157**:153-  
385 161.

386

387 Toma C, Martínez Espinosa E, Song T, Miliwebsky E, Chinen I, Iyoda S, Iwanaga M and  
388 Rivas M (2004) Distribution of Putative Adhesins in Different Seropathotypes of Shiga  
389 Toxin-Producing *Escherichia coli*. *J Clin Microb* **42**:4937-4946.

390

391 Torres AG, Blanco M, Valenzuela P, Slater TM, Patel SD, Dahbi G, López C, Barriga XF,  
392 Blanco JE, Gomes TAT, Vidal R, Blanco J (2009) Genes related to long polar fimbriae of  
393 pathogenic *Escherichia coli* strains as reliable markers to identify virulent isolates. *J Clin*  
394 *Microbiol* **47**:2442-2451.

395

396 Torres AG, Slater TM, Patel SD, Popov VL, Arenas-Hernández MMP (2008) Contribution of  
397 the Ler- and H-NS-regulated long polar fimbriae of *Escherichia coli* O157:H7 during binding  
398 to tissue-cultured cells. *Infect Immun* **76**:5062-5071.

399

400 Touchon M, Hoede C, Tenailon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S,  
401 Bouchier C, Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller S, Danchin A, Diard  
402 M, Dossat C, Karoui ME, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguéneq  
403 C, Lescat M, Mangenot S, Martinez-Jéhanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon

404 C, Rouy Z, Ruf CS, Schneider D, Turret J, Vacherie B, Vallenet D, Médigue C, Rocha EPC,  
405 Denamur E (2009) Organised genome dynamics in the *Escherichia coli* species results in  
406 highly diverse adaptive paths. *PLoS Genet* **5**:e1000344.

407

408 Tóth I, Schmidt H, Kardos G, Lancz Z, Kreuzburg K, Damjanova I, Pászti J, Beutin L, Nagy  
409 B (2009) Virulence genes and molecular typing of different groups of *Escherichia coli* O157  
410 strains in cattle. *Appl Environ Microbiol* **75**:6282-6291.

411

412 **Table 1. List of strains used in this study and results of the PCR scanning of the flanking regions of the *lpf2* operon.**

Strain	Serotype	Phylogenetic group	<i>bgIG-phoU</i>	<i>phoU-pstB</i>	<i>pstB-pstA</i>	<i>pstA-pstC</i>	<i>pstC-pstS</i>	<i>pstS-lpfD</i>	<i>lpfA-glmS</i>	<i>glmS-glmU</i>	<i>glmU-atpC</i>	Reference
T16	O157:H43	B1 <sup>c</sup>	+	+	+	+	+	+	+	+	+	(Tóth et al., 2009)
T22	O157:H43	B1 <sup>c</sup>	+	+	+	+	+	+	+	+	+	(Tóth et al., 2009)
T34	O157:H43	B1 <sup>c</sup>	+	+	+	+	+	+	+	+	+	(Tóth et al., 2009)
T49	O157(rough):H9	B1 <sup>c</sup>	+	+	+	+	+	+	+	+	+	(Tóth et al., 2009)
T50	O157(rough):H37r	B1 <sup>c</sup>	+	+	+	+	+	+	+	+	+	(Tóth et al., 2009)
B47	O157:NM	B1 <sup>c</sup>	+	+	+	+	+	+	+	+	+	(Tóth et al., 2009)
B54	O157(rough):H12	A <sup>c</sup>	- <sup>a</sup>	+	- <sup>a</sup>	+	+	+	+	+	+	(Tóth et al., 2009)
ECOR7	O85:HN	A	+	+	+	+	+	+	+	+	+	(Ochman & Selander, 1984)
ECOR23	O86:H43	A	+	+	- <sup>a</sup>	+	-	+	+	+	+	(Ochman & Selander, 1984)
ECOR26	O104:H21	B1	+	+	+	+	+	+	-	+	+	(Ochman & Selander, 1984)
ECOR30	O113:H21	B1	+	+	+	+	+	+	+	+	+	(Ochman & Selander, 1984)
ECOR32	O7:H21	B1	+	+	+	+	+	+	+	+	+	(Ochman & Selander, 1984)
ECOR33	O7:H21	B1	+	+	+	+	+	+	+	+	+	(Ochman & Selander, 1984)
ECOR34	O88:NM	B1	+	+	-	+	+	+	+	+	+	(Ochman & Selander, 1984)
ECOR36	O79:H25	D	+	+	+	+	+	+	+	+	+	(Ochman & Selander, 1984)
ECOR57	ON:NM	B2	- <sup>b</sup>	- <sup>b</sup>	+	+	- <sup>b</sup>	(Ochman & Selander, 1984)				
ECOR58	O112:H8	B1	+	+	+	+	+	+	+	+	+	(Ochman & Selander, 1984)
ECOR67	O4:H43	B1	+	+	+	+	+	+	+	+	+	(Ochman & Selander, 1984)
ECOR69	ON:NM	B1	+	+	+	+	+	+	+	+	+	(Ochman & Selander, 1984)
ECOR72	O144:H8	B1	+	+	+	+	+	+	+	+	+	(Ochman & Selander, 1984)
187/06 (22)	O136:H12	B2	+	+	- <sup>a</sup>	+	+	+	+	+	+	this study
C600	K12	A	-	+	+	+	-	+	+	+	+	(Appleyard, 1954)

413

414 <sup>a</sup>These strains yielded consistently longer product with the given primer pair.

415 <sup>b</sup> Strain ECOR57 yielded unspecific or weak products in all these reactions.

416 <sup>c</sup> The phylogenetic group of these strains was determined in Sváb & Tóth, 2012.

417 Abbreviations:

418 HN: H antigen non-typeable

419 NM: non-motile

420 ON: O antigen non-typeable

421

422 **Table 2. Primers used for the amplification of regions flanking the *lpf2* operon.**

Primer	Sequence (5'->3')	Genes amplified	Genbank number	Position amplified	Annealing temperature (°C)	Reference
bgIGfw	CCCAAGCGCTCCTGCGCTAAA	<i>bgIG-phoU</i>	CU928160.2	3977599-3978492	60	this study
phoUrev phoUfw	TCTGGAGTCGCTGGGCCGTC CGATCACGCGCTTCGCCAGA	<i>phoU-pstB</i>	CU928160.2	3978707-3979162	59	this study this study
pstBrev pstBfw	GGTATCGCCATTCGCCCCGGA GGACGGCCCCGATAAACGCCG	<i>pstB-pstA</i>	CU928160.2	3979526-3979912	60	this study this study
pstArev pstAfw	CAGCCGATCGCCAACCTGCC AGCCAGAACAGGCCGAAGGC	<i>pstA-pstC</i>	CU928160.2	3980508-3980919	59	this study this study
pstCrev pstCfw	ATCGGCGGCATCATGCTGGG ACCGTAGATCGGCACCAGCG	<i>pstC-pstS</i>	CU928160.2	3981370-3981924	58	this study this study
pstSrev pstSfw	CCAGAAAGGCGAAGATGCATGGC CAGACAGCGGCGCGTCAGAG	<i>pstS-lpfD</i>	CU928160.2	3982472-3983233	58	this study this study
lpfDrev lpfAfw	TGCTACCGAACCCAATACGGACAA TGTCGACAATTTACCGACGAAGT G	<i>lpfA-glmS</i>	CU928160.2	3987849-3988769	58	this study this study
glmSrev glmSfw	GCTGCCGAGCCGTATTGAGCA CGTGTGTCGCCAGCGAGTA	<i>glmS-glmU</i>	CU928160.2	3989854-3990519	58	this study this study
glmUrev glmUfw	GGCGATGCGGAAATTGGCGA AGCAGATCGCCGCCGTGAC	<i>glmU-atpC</i>	CU928160.2	3991436-3992119	59	this study this study
atpCrev lpfA F	ACGAAGCGCGAGCCATGGAA ACCGCTATCGATGCTGAAGG	<i>lpfB-lpfA</i>	AY057066	678-1349	63	this study (Ideses et al., 2005)

lpfB R	GCGCAACATCTTCGGGAATA					
lpfC F2	CGCCGGGTTAGAAATAGATA	<i>lpfD-lpfC</i>	AY057066	3658-4421	63	(Ideses et al., 2005)
lpfD R2	TGCCTGGTTTATTTTTGACGTA					
lpfA_inside_fw <sup>a</sup>	TCGACAGTAAATTGTGAATC	Part of <i>lpfA</i>	AY057066	233-790	50	this study
lpfA_inside_rev <sup>a</sup>	GAAGCGTAATATTATAGGCG					

---

423

424 <sup>a</sup> Primers used in RT-PCR for the detection of *lpfA* expression.

425

426 **Table 3. *Escherichia coli* strains with whole genomes in GenBank containing continuous *lpf2* operons with high homology to the *lpf2-1* of**  
 427 **strain T22.**

Strain	Serotype	Pathotype	SNPs in the Lpf operon	Genbank number	Reference
SE11		commensal	6	AP009240.1	(Oshima et al., 2008)
11128	O111:NM	EHEC	7	AP010960.1	(Ogura et al., 2009)
55989		EAEC	8	CU928145.2	(Touchon et al., 2009)
KO11		commensal	9	CP002516.1	unpublished, JGI Project ID: 4085738
W		commensal	9	CP002185.1	(Archer et al., 2011)
11368	O26:H11	EHEC	9	AP010953.1	(Ogura et al., 2009)
IA11		commensal	10	CU928160.2	(Touchon et al., 2009)
E24377A		ETEC	11	CP000800.1	(Rasko et al., 2008)

428

429 **FIGURE LEGENDS**

430

431 **Figure 1. Schematic representation of the genetic region from *E. coli* strain T22 containing *lpf2*.** The relative length of the arrows is  
432 proportional to the relative length of the genes. Arrows representing genes from the same functional cluster are filled with the same pattern. The  
433 *pst* cluster encodes a phosphate ABC transporter, *phoU* encodes a phosphate transport regulator, and the product of *bgIG* is a putative  
434 transcriptional regulator. The *glm* cluster encodes an N-acetyl glucosamine-1-phosphate uridylyltransferase, and *atpC* encodes the epsilon subunit  
435 of ATP synthase.

436

437 **Figure 2. Comparative chart of *in vitro* bacterial adhesion assays to Caco-2, Hep-2 and T84 cell lines.** The results are expressed as the  
438 percentage of cell-associated bacteria from the original inoculum [(final CFU/ml / initial CFU/ml) x 100] and are the means  $\pm$  the standard error  
439 of at least two independent experiments in triplicate wells.

440

441

442 **Figure S1. Sequence comparison of the *lpf2* operons of atypical bovine *Escherichia coli* O157 strains.** The yellow arrows indicate the ORFs  
443 of the individual *lpf* genes. Point mutations are marked in red. The sequences of the *lpf2* operons were deposited under the following accession  
444 numbers: *E. coli* B47 (O157:NM) **KC207119**, *E. coli* B54 (O157:H12) **KC207120**, *E. coli* T16 (O157:H43) **KC207121**, *E. coli* T34 (O157:H9)  
445 **KC207122**, *E. coli* T49 (O157:H37) **KC207123**, *E. coli* T50 (O157:H43) **KC207124**

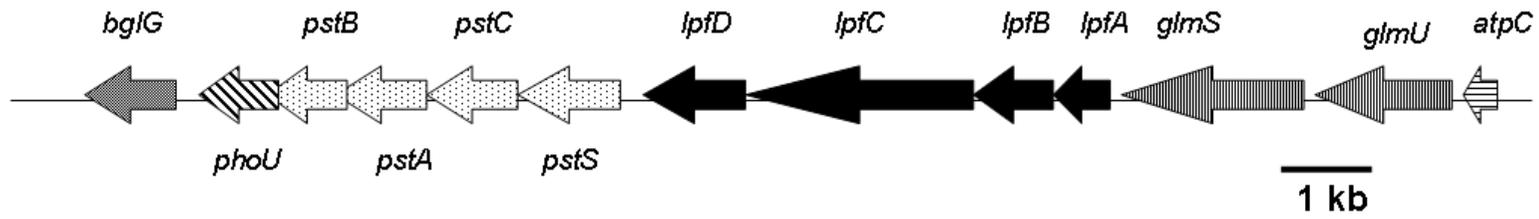
446

447

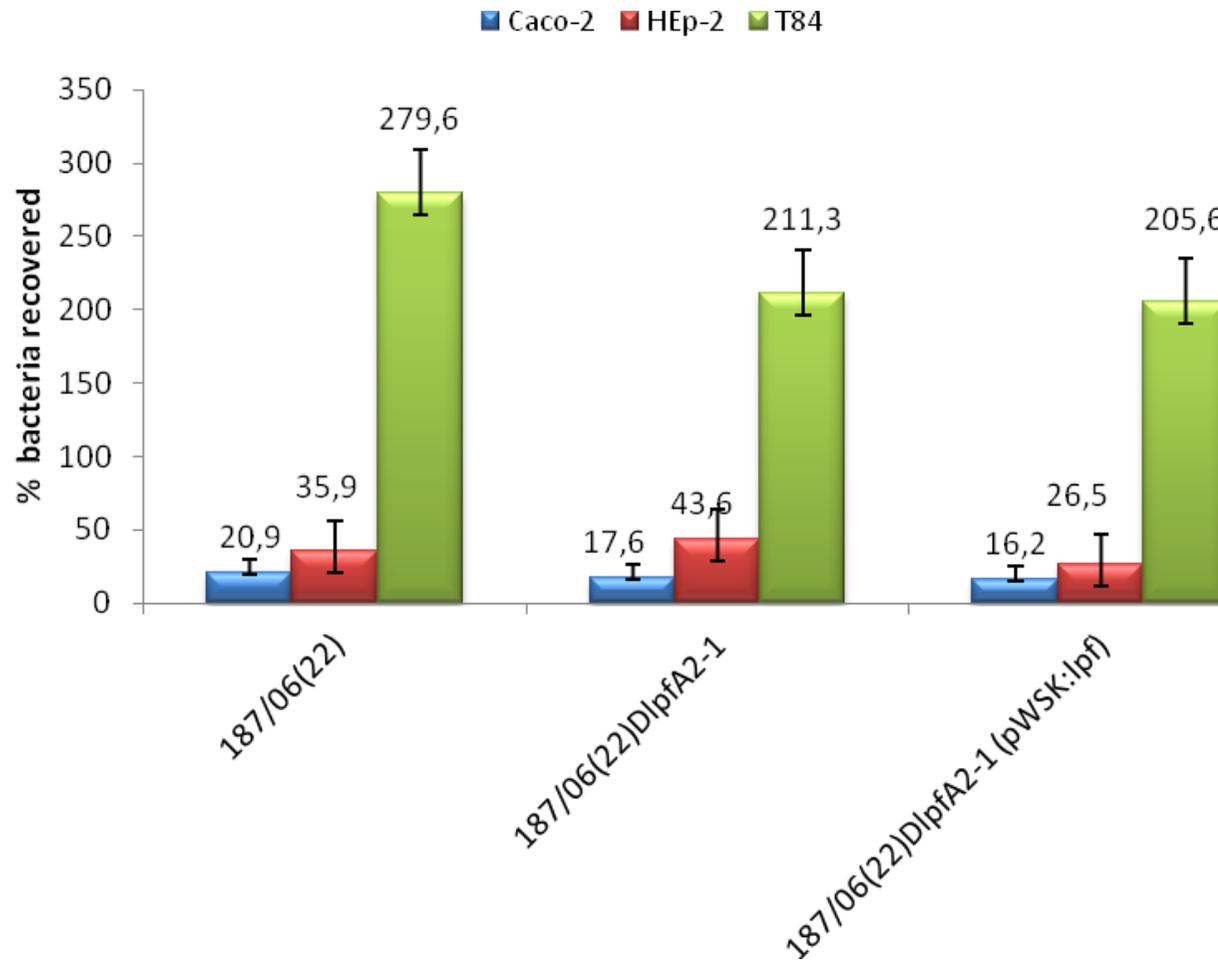
448 **FIGURES**

449

450 **Figure 1. Schematic representation of the genetic region from *E. coli* strain T22 containing *lpf2*.**



452 **Figure 2. Comparative chart of *in vitro* bacterial adhesion assays to Caco-2, Hep-2 and T84 cell lines.**



453

