1	AEM00699-10 Revision
2	
3	Animal and human pathogenic Escherichia coli strains share common genetic
4	backgrounds
5	
6	
7	Olivier Clermont ¹ , Maiwenn Olier ² , Claire Hoede ¹ , Laure Diancourt ³ , Sylvain Brisse ³ ,
8	Monique Keroudean ² , Jérémy Glodt ¹ , Bertrand Picard ⁴ , Eric Oswald ^{2, 5} , Erick Denamur ¹
9	
10	
11	¹ UMR722, INSERM and Université Paris 7, Site Xavier Bichat, 75018 Paris, France
12	² INRA, UMR 1225, 31076 Toulouse, France
13	³ Genotyping of Pathogens and Public Health, Institut Pasteur, 75724 Paris, France
14	⁴ UMR722, INSERM and Université Paris 13, Site Xavier Bichat, 75018 Paris, France
15	⁵ Laboratoire de Bactériologie-Hygiène, CHU de Toulouse, Institut Fédératif de Biologie,
16	31059 Toulouse, France
17	
18	Corresponding author: Erick Denamur, INSERM U722, Site Xavier Bichat, 16 rue Henri
19	Huchard, 75018, Paris, France. Tel: 33 1 57 27 75 34. Email: erick.denamur@inserm.fr
20	
21	
22	Key words: Escherichia coli, phylogeny, host specificity, pathogenic
23	
24	Running title: Animal and human pathogenic E. coli relationships

- 1 Abstract
- 2

3 Escherichia coli is a versatile species encompassing both commensals of the digestive 4 tracts of many vertebrates, including humans, and pathogenic strains causing various intra-5 and extraintestinal infections. Despite extensive gene flow between strains, the E. coli species 6 has a globally clonal population structure, consisting of distinct phylogenetic groups. Little is 7 known about the relationships between phylogenetic groups and host specificity. We therefore 8 used multi-locus sequence typing (MLST) to investigate phylogenetic relationships and 9 evaluated the virulence gene content of 35 E. coli strains representative of the diverse diseases 10 encountered in domestic animals and compared these strains with a panel of 100 human 11 pathogenic and 99 non human and human commensal strains representative of the 12 phylogenetic and pathovar diversity of this species. A global factorial analysis of 13 correspondence indicated that extraintestinal infections were caused mostly by phylogenetic 14 group B2 strains, whereas intraintestinal infections were caused mostly by phylogenetic group 15 A/B1/E strains, with each group having specific virulence factors. It was not possible to 16 distinguish between strains of human and animal origin. A detailed analysis showed that some 17 pathogenic human and animal strains belonged to the same clonal complexes, with a number 18 of virulence genes in common. However, a set of specific adhesins was identified in animal 19 non-B2 group strains whatever the *E. coli* pathotype. In conclusion, human and animal 20 pathogenic strains share common genetic backgrounds, but in non-B2 phylogenetic group 21 strains, seem to have a different set of adhesins according to their origin.

1 Introduction

2

3 Escherichia coli is one of the most versatile bacterial species. It alternates between its 4 primary habitat, the gut of vertebrates, where it lives as a commensal (50), and its secondary 5 habitat, water and sediment (49). It may also function as an intra- and extraintestinal pathogen 6 in humans and many other animal species (30). This diversity of lifestyles is achieved through 7 a high degree of genome plasticity, with gains, through horizontal transfer, and losses of 8 genes (45, 51). This species has a core genome of less than 2,000 genes, but the total number 9 of genes in the species exceeds 10,000 (45, 51). Thus, the diverse phenotypes observed result 10 principally from a large number of different gene combinations. Despite the high degree of 11 gene flow, the structure of this species remains mostly clonal (51), with the clear delineation 12 of at least six principal phylogenetic groups (A, B1, B2, D, E and F) (25, 50). It has been 13 shown that genetic background plays a role in the acquisition, retention and expression of 14 foreign DNA (14).

15 It has been suggested that some group B1 strains with the hly gene may be specific for 16 animals (15) and that some group B2 strains of the O81 serogroup may be specific for 17 humans (11). However, little is known about the relationships between phylogenetic groups 18 and host specificity. In this context, the extent to which bacterial strains from infected humans 19 and animals are phylogenetically related is unclear. Several studies based on serotyping, 20 multilocus enzyme electrophoresis, outer membrane protein profiles, pulsed-field gel 21 electrophoresis, ribotyping, randomly amplified polymorphic DNA, phylogenetic group 22 affiliation and virulence gene content have shown close relationships between human and 23 animal isolates (1, 6, 7, 16, 19, 20, 27, 29, 34, 35, 39, 44, 53). More recently, multilocus 24 sequence typing (MLST) has been used to study the phylogenetic relationships between strains in more detail. These studies have focused principally on extraintestinal pathogenic E. 25

coli (ExPEC) [including avian pathogenic *E. coli* (APEC) in particular] (37-39) and enterohemorrhagic *E. coli* (EHEC) (17, 40) strains. A large body of data has accumulated, but these data are scattered and difficult to compare, as these studies used different typing approaches and non-redundant sets of strains. Consequently, no overview of the global relationships between animal and human pathogenic strains in the framework of the phylogeny of the whole *E. coli* species has yet been produced.

7 The aim of this work was to study the phylogenetic relationships, by MLST, and 8 virulence gene content of 35 *E. coli* strains representative of the diverse diseases encountered 9 in domestic animals, comparing these strains with a panel of 100 human pathogenic and 99 10 non human and human commensal strains representative of the phylogenetic diversity of the 11 species and including well characterized archetypal strains.

12

13 Materials and Methods

14

15 Bacterial strains. We studied 234 E. coli strains and one strain of Escherichia fergusonii, the 16 closest relative of E. coli (31) (Table S1). Five groups of E. coli strains were represented: (i) a 17 panel of 35 pathogenic animal strains representative of the various diseases encountered in 18 domestic animal species (from 8 birds and 27 mammals) comprising 15 ExPEC/APEC and 20 19 intraintestinal pathogenic E. coli (InPEC) [8 enterotoxigenic E. coli (ETEC), 4 20 enteropathogenic E. coli (EPEC), 5 Shiga toxin-producing E. coli (STEC)/EHEC and 3 21 unclassified InPEC] strains (Table 1), (ii) a panel of 92 pathogenic human strains comprising 22 43 ExPEC [29, 8 and 6 involved in urinary tract infection (UTI), newborn meningitis (NBM), septicemia and miscellaneous infections, respectively] and 49 InPEC [7 ETEC, 6 EPEC, 8 23 24 EHEC, 10 enteroaggregative E. coli (EAEC), 1 enteroinvasive E. coli (EIEC), 16 diffusely adherent E. coli (DAEC) and 1 unclassified InPEC], (iii) 45 non human mammalian 25

commensal strains, (iv) 54 human commensal strains and (v) 8 human InPEC strains for which the complete genome was available (42, 45), these strains being typed *in silico* in the present work. The strains from groups (ii), (iii) and (iv) in this list originated mostly from three published collections (14, 32, 41) and may be considered representative of the phylogenetic and pathovar diversity of *E. coli*. They encompass archetypal strains for various diseases, and complete genome sequences are available for some.

7

8 Virulence factor (VF) screening. We tested for the presence of virulence factors involved in 9 extraintestinal (neuC, kpsE, sfa/foc, iroN, aer, iha, papC, papGI, papGII, papGIII, hly, cnf1, 10 hra, sat, ire, usp, chromosomal ompT, ibeA, malX, irp2, fyuA and traT) (12, 26) and 11 intraintestinal [afaD, ipaH, stx1, stx2, eltB (LT), estA (ST), bfpA, eae, aaiC and aatA] (Table 12 S2) infections by PCR, as previously described (14). We also used PCR, as previously 13 described (2, 4, 13, 18, 24), to check for the presence of adhesins classically associated with 14 animal-specific pathogenic strains of E. coli: K99 (fanA), K88 (faeG), F17 (f17A), F18 (fedA) 15 and Afr2 (afr2G).

16

PCR O-typing. O-type was determined by an allele-specific PCR (10) using the primers given in Table S3. We assessed 28 O-types with this assay, based on the O types previously determined by the classical serological method for the other strains of the collection.

20

MLST. MLST was performed with partial *dinB*, *icdA*, *pabB*, *polB*, *putP*, *trpA*, *trpB*, and *uidA* sequences (25). Allele sequences and sequence types (STs) are available from Institut Pasteur's MLST website, at <u>www.pasteur.fr/mlst</u>. Phylogenetic analysis was performed with the concatenated sequences of the eight genes, by the maximum likelihood method, as implemented in the PHYML program (23), with *E. fergusonii* as the outgroup.

2 Factorial analysis of correspondence (FAC). FAC was used to describe associations 3 between the different data sets. FAC uses a covariance matrix based on Chi squared distances 4 (22). This computation method determines a plane defined by two principal axes of the 5 analysis. The first axis (F1) accounts for most of the variance, and the second axis (F2), 6 orthogonal to F1, accounts for the largest part of the variance not accounted for by F1. FAC 7 was conducted with SPAD.N 4.5 software (Cisia, Saint Mandé, France), based on a two-way 8 table. This table had 234 rows, one for each E. coli strain, and 35 columns, corresponding to 9 35 variables: human/animal origin, commensal, ExPEC and InPEC characters, seven 10 phylogenetic groups corresponding to A, B1, B2, C, D, E, F and ungrouped (UG) strains 11 according to the MLST data (14, 25), and 23 VFs (neuC, sfa/foc, iroN, papC, papGI, papGII, papGIII, hlyC, cnf1, hra, fyuA, "animal adhesins" (AnAd), afaD, ipaH, stx1, stx2, estA, eltB, 12 13 *bfpA*, *eae*, *aaiC* and *aatA*). The data in this table were attributed a binary code: "1" for present 14 and "0" for absent. The loading score for each variable on the plane of the variables (factors 15 F1 and F2, respectively) can be inferred from the coordinates of the X and Y variables on the 16 F1/F2 plane. Moreover, the data corresponding to active variables were calculated directly in 17 the FAC, whereas the illustrative variables were only projected on the plane.

18

- 19 **Results and Discussion**
- 20

21 Multidimensional analysis

22

We assessed the global relationships between the phylogenetic groups, the VF content and origin (human versus animal and commensal versus pathogenic) of the strains, by carrying out a FAC with phylogenetic groups and VFs as active variables and the

1 animal/human, commensal, ExPEC and InPEC variables as illustrative variables. Strains were 2 assigned to phylogenetic groups on the basis of MLST data. Seven groups were considered: 3 the six previously recognized groups (A, B1, B2, D, E and F) and the C group plus ungrouped 4 strains (see below for the definition of the C group and ungrouped strains). On the F1/F2 5 plane, which accounted for 29.37 % of the total variance, the variables *papGI*, *papGIII*, *cnf1*, 6 sfa/sfoc, hlvC, iroN, hra, papC, neuC, B2 group and ExPEC were projected onto the positive 7 values of the F1 axis, whereas the variables stx2, E group, eae, stx1, bfpA, eltB, estA, B1 8 group, InPEC, AnAd, ipaH, aatA and A group were projected onto the negative values of this 9 axis. The variables F group, papGII, D group, aer, afaD and aaiC were projected onto the 10 negative values of the F2 axis. The variables human, animal and commensal were grouped 11 around the origin of the axes and it was therefore not possible to differentiate between them in 12 this FAC (Fig. 1). Most of the ExPEC had positive F1 coordinates (Fig. 2A), whereas the 13 InPEC strains had negative coordinates on this axis (Fig. 2B). Moreover, it was not possible 14 to distinguish between ExPEC and InPEC strains of human and animal origin in this FAC 15 (Fig. 2).

16 This global analysis suggests that extraintestinal infections are caused principally by 17 B2 strains, which have many extraintestinal VFs, whereas intraintestinal infections are caused 18 mostly by A/B1/E strains with specific intestinal VFs. It was not possible to distinguish 19 between strains of human and animal origin.

20

21 Fine-scale phylogenetic analysis

22

We investigated the relationships between human and animal pathogenic strains in more detail, at the clonal level and within the species as a whole, by carrying out MLST analysis on the 234 strains. Three main MLST schemes are currently available for *E. coli* (25, 46, 52), using each a different combination of genes. The data are usually studied in two ways, with the alleles at different loci providing an allelic profile, which defines the ST. No weighting is given to take into account the number of nucleotide differences between the alleles. Alternatively, nucleotide sequences may be used for phylogenetic reconstructions (50). The results obtained with these three schemes are highly correlated (21), suggesting that the clonal structure of the species is robust.

7 We chose to use the set of genes described in a previous study (25), as we have 8 previously used this MLST scheme to characterize a unique set of strains representative of the 9 phylogenetic and lifestyle diversity of the species. A considerable amount of data is available 10 for these strains. The phylogenetic tree reconstructed from the concatenated sequences (Fig. 11 3) showed the major phylogenetic groups previously described (50): A, B1, B2, D, E, F. An 12 additional group, closely related to the B1 group but identified as A by PCR triplex 13 phylogrouping (9), was also identified and called C (14, 36). Only six strains (ECOR31, 14 ECOR42, 126A, DAEC5, DEC9a and 101-1) were not included in these groups and were 15 classified as ungrouped. The F and B2 groups are the most basal, having emerged first, whereas the A and B1/C groups diverged more recently. Furthermore, a clear genetic 16 17 structure was identified within the B2 phylogenetic group, with at least nine subgroups (I to 18 IX) in addition to the EPEC1 cluster commonly represented by the O127:H6 strain E2348/69 19 (32). The A and B1 groups contained three (I to III) and five (I to V) subgroups, respectively 20 (Fig. 3). However, with the exception of subgroup II of group A, which was delineated by a 21 high bootstrap value (84%) and encompassed an A₀ genotype with the PCR triplex typing 22 method (9, 21), these subgroups were not particularly robust phylogenetically. A minimum 23 spanning tree analysis based on allelic profile data gave similar results for the shallow 24 phylogenetic grouping (data not shown).

With this phylogenetic approach, we clearly identified human and animal strains belonging to the same phylogenetic subgroups or clonal complexes.

2

3 Both human and animal pathogenic strains were found in all but three of the 4 subgroups of the B2 phylogenetic group (I, V and VIII). The most highly represented 5 subgroup was subgroup IX, which corresponds to ST95 of the Achtman scheme (52), ST29 of 6 the Whittam scheme (40), and the B2-1 group of a previous study (38). This subgroup 7 encompasses the APEC strains of serogroup O1, O2 and O18 (28, 37, 38) and the archetypal 8 human strains UTI89 (5), a strain causing urinary tract infection, and RS218, an isolate from a 9 neonate with meningitis (54), both of serotype O18:K1:H7. Non-human mammalian ExPEC 10 strains were also found in subgroup II (O2-type strain, archetypal UTI human strain CFT073 11 of O6-type), subgroup IV (O2-type strain, archetypal UTI human strain IAI74 of O2-type) 12 and subgroup VII (O75-type strain, archetypal human UTI strain IH11128 of O75-type). 13 Furthermore, strains of serotype O6:H31, which have been reported in dog urinary tract 14 infections (7, 27), belong to subgroup III (archetypal human UTI strain 536 of O6-type) and 15 O4:H5 dog strains belong to subgroup VI (archetypal human UTI strain J96 of O4-type) (32) 16 (data not shown). As expected, no pathogenic animal strain belonged to subgroup VIII, as this 17 clone has been described as strictly commensal in humans (11). Similarly only human EPEC 18 strains testing positive for *bfpA* by PCR were present in the EPEC 1 cluster.

Non B2 ExPEC animal strains of the O78-type belong to phylogenetic groups D (one APEC strain) and C (2 APEC and 2 bovine ExPEC strains). Group D is diverse, and no close relationship was identified between group D strains and human strains, although numerous human ExPEC strains belong to this group (3, 43). Conversely, the group C is clonal. Thus, the numerous APEC strains of the O78-type assigned to phylogroup A by the triplex PCR method (16, 29) probably belong to this clone. One of the group C strains tested, the bovine ExPEC strain 1404, which carries the Vir plasmid, has been shown to be closely related to

avian and human ExPEC strains, as it belongs to esterase electrophoretic type 2 (6). The
group C strains in our collection also include one human ExPEC strain, ECOR72, and animal
and human InPEC strains. A case of neonatal meningitis due to a group C strain has recently
been reported (36). This strain was shown to be virulent in a mouse model of septicemia (43).
One ExPEC strain from pig and another from sheep, these strains belonging to the B1
subgroups III and V, respectively, clustered together with human InPEC strains.

7 Animal InPEC strains were found to belong to phylogenetic groups A, B1, C and E. 8 Group E, which includes human and animal EHEC O157:H7 strains, is clonal, whereas 9 groups A and B1 are more diverse, with at least three (I to III) and five (I to V) subgroups, 10 respectively (Fig. 3). All animal InPEC (ETEC and EHEC) strains of phylogenetic group A 11 belong to subgroup I, which also contains human InPEC (DAEC and EAEC with the 12 archetypal EAEC strain JM221) and ExPEC strains (Fig. 3). Four of the five B1 subgroups 13 include animal InPEC strains. Within the B1 subgroups, bovine O26-type EHEC and EPEC, 14 together with O45b-type and O15-type EPEC strains, are closely related to human O26-type 15 and O111-type EHEC strains. The swine ETEC 987 strain belongs to B1 subgroup III, which 16 encompasses the human EAEC, EPEC and ETEC strains. The rabbit EPEC strain E22 (O103-17 type) is very closely related to the human O103-type and O111-type EHEC strains and to the 18 EPEC 2 strain B171 (45), in B1 subgroup V. Animal ETEC and EHAC strains were found in 19 group C, together with human InPEC and animal and human ExPEC strains, as stated above 20 (Table 1).

21

22 Fine-scale VF pattern analysis

23

Animal ExPEC strains have a variable pattern of extraintestinal VFs, whether considered individually or in the pathogenicity island (PAI) context, even within a clonal

1 lineage (Table 1). Different patterns were also observed between human and animal strains of 2 the same lineage (for example, within B2 subgroups II, VII and IX; Tables 1 and S1). 3 Furthermore, no particular pattern of VFs was found to be specific to a particular type of 4 extraintestinal disease (e.g. UTI, NBM, septicemia or avian colibacillosis). The intraintestinal 5 VFs studied were pathotype-specific and therefore did not distinguish between human and 6 animal strains. In a very small number of strains in our collection, we did not find the 7 expected pathotype-defining VFs (ST, LT, Bfp). However, in these cases, the VFs were 8 plasmid-encoded and the plasmid was probably lost during the many subcultures of the strains 9 (Table 1).

10 By contrast, adhesins, which are usually described in animal strains (see above), were 11 found in 17 animal strains but only four human strains: the F111 adhesin in 1 DAEC strain, 1 12 EAEC strain and 1 ExPEC strain, and the F17c/GAF adhesin in a ExPEC strain (Tables 1 and 13 S1). The expression of the F17c/GAF adhesin, which is specific for terminal N-acetyl-D-14 glucosamine and blood group M antigen, by human strains (particularly those isolated from cases of urinary tract infection) is well documented, although the reported incidence is low 15 16 (47). By contrast, to our knowledge, the presence of genes encoding F111 adhesins has not 17 previously been reported in an E. coli strain pathogenic in humans. As suggested by the FAC 18 (Fig. 1), these adhesins were found almost exclusively in non-B2 strains, as only three of the 19 strains concerned belonged to phylogenetic group B2.

20

21 Concluding remarks

22

Our data for a representative set of animal pathogenic strains analyzed within the context of the overall phylogeny of the *E. coli* species clearly show that human and animal pathogenic strains share common genetic backgrounds, with a large overlap between the two populations. Human and animal strains causing the same disease in different host share a
 common pool of virulence genes, but a set of specific adhesins was identified in animal non B2 strains whatever the *E. coli* pathotype.

4 It has been suggested that, on many occasions, strains derived from a common recent 5 ancestor have become specialized for a particular host through subtle genetic changes (48). 6 The recent identification of genes or single nucleotide polymorphisms in O157:H7 EHEC 7 strains specific for humans or cattle (8, 53) and of specific expression patterns of genes from 8 B2 phylogenetic group APEC and human urinary tract infection strains in chicken and UTI 9 mouse models (55) is consistent with this hypothesis. Complete genome sequences for many animal and human pathogenic strains, generated by "next-generation" sequencing 10 11 technologies (33), will facilitate identification of the genetic elements involved in host 12 specialization.

2 Acknowledgments

- 3
- 4 ED was supported in part by the "Fondation pour la Recherche Médicale".

Strain ID	Host	Country origin	Pathogenicity	Phylogenetic group and subgroup	O-type	neuC	kpsE	sfa/foc	iroN	aer	papC	papGI	papG II	papG III	hlyC	cnfI	hra	fyuA	"Animal" adhesin	afaD	ipaH	stxI	stx2	eltB (LT)	estA (ST)	bfpA	eae	aaiC	aatA
25KH9	Bos taurus	Belgium	ETEC	ΑI	101	-	-	-	+	-	+	-	-	-	-	-	+	-	F17a	-	-	-	-	-	-	-	-	-	-
S1191	Sus scrofa	USA	STEC	ΑI	139	-	-	-	-	-	-	-	_	-	-	-	-	-	F18	-	-	-	+	-	+	-	-	-	-
510	Bos taurus	Belgium	ETEC	ΑI	101	-	-	-	-	-	-	-	-	-	-	-	+	-	K99	-	-	-	-	-	+	-	-	-	-
431	Sus scrofa	USA	ETEC	ΑI	101	-	-	-	-	-	-	-	-	-	-	-	+	-	K99	-	-	-	-	-	+	-	-	-	-
262KH89	Bos taurus	Belgium	Diarrhea	ΑI	26	-	-	-	-	-	-	-	_	-	+	-	-	-	K88	-	-	-	-	-	-	-	-	+	-
255/1-1	Bos taurus	France	EHEC	ΑI	Unknown	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	-
126A	Bos taurus	Belgium	ETEC	UG	8	-	-	-	-	-	-	-	-	-	-	-	+	-	K99	-	-	-	-	-	+	-	-	-	-
86-1390	Sus scrofa	Canada	EPEC	B1 I	45b	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-	-
193	Bos taurus	USA	EHEC	B1 I	26	-		-	-	+	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	+	-	-
C/15333	Bos taurus	Northern Ireland	EPEC	B1 I	26	_	-	_	-	_	-	_	-	_	+	_	-	+	-	_	_	-	-	_	-	_	+	_	_
RDEC-1	Oryctolagus cuniculus	USA	EPEC	B1 I	15	-	_	-	-	+	-	-	-	-	-	-	-	+	-	_	_	_	_	_	_	_	+	_	_
5131	Sus scrofa	Canada	ExPEC	B1 III	115	-	-	+	+	+	+	_	_	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
987	Sus scrofa	USA	ETEC	B1 III	9	-	-	-	-	-	-	-	-	-	_	-	-	-	K99	-	-	-	_	_	+	_	-	-	_
31A	Bos taurus	France	Diarrhea	B1 IV	153	-	-	-	-	+	+	-	_	-	_	-	+	+	F17c/GAF	_	-	-	_	_	-	-	-	-	-
211	Bos taurus	Belgium	ETEC	B1 V	Rough	+	+	-	-	-	-	_	_	-	_	-	+	-	-	-	-	-	_	_	-	_	-	-	_
S5	Ovis aries	UK	ExPEC	B1 V	15	+	+	-	-	-	-	_	_	-	+	-	-	+	F17b	-	-	-	_	_	-	_	-	-	_
E22	Oryctolagus cuniculus	France	EPEC	B1 V	103	-	-	-	-	-	-	-	-	-	-	-	-	-	AFR2	_	_	_	_	_	_	_	+	_	_
111KH86	Bos taurus	Belgium	Diarrhea	B1 V	Unknown	-	-	-	-	-	-	-	-	-	-	-	+	-	F111	-	-	-	-	-	-	-	-	-	-
789	Gallus gallus	Israël	APEC	С	78	_	-	+	+	+	_	-	-	_	_	-	_	+	-	-	_	_	-	_	_	_	_	-	_
BEN 0265	Gallus gallus	France	APEC	С	78	-	-	-	+	+	-	-	-	-	_	-	-	+	-	-	-	-	-	_	-	-	-	-	-
1404	Bos taurus	France	ExPEC	С	78	-	-	-	-	-	-	-	_	_	+	-	-	+	F17b	-	-	-	-	-	-	-	-	-	-
239KH89	Bos taurus	Belgium	ExPEC	С	Unknown	-	-	-	-	+	+	-	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
DEC7a	Sus scrofa	USA	ETEC	С	157	-	-	-	-	-	-	-	-	-	+	-	-	+	K88	-	-	-	-	+	+	-	-	-	-

1 Table 1. Main characteristics of the 35 pathogenic animal *E. coli* strains studied

248/1-2	Bos taurus	France	EHEC	С	Unknown	-	-	-	+	-	+	-	-	-	-	-	+	+	-	+	-	+	-	-	-	-	-	-	-
G7	Sus scrofa	UK	ETEC	С	8	-	-	-	-	-	-	-	-	-	+	-	-	+	K88	-	-	-	-	+	-	-	-	-	-
DEC4a	Bos taurus	Argentina	EHEC	Е	157	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	-	-
BEN 1189	Gallus gallus	Belgium	APEC	D	78	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BM2-1	Bos taurus	France	ExPEC	B2 II	2	-	-	+	+	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-
M623	Sus scrofa	Spain	ExPEC	B2 IV	2	+	+	-	+	-	+	-	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
28C	Sus scrofa	Spain	ExPEC	B2 VII	75	-	-	+	+	-	+	-	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
APECO1	Gallus gallus	USA	APEC	B2 IX	1	+	+	-	+	+	+	-	+	-	_	-	-	+	-	-	-	-	-	-	-	-	-	-	-
BEN 0139	Meleagridis gallopavo	France	APEC	B2 IX	2	+	+	-	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-
BEN 2908	Gallus gallus	France	APEC	B2 IX	2	+	+	+	+	+	-	-	-	-	-	-	+	+	F111	-	-	-	-	-	-	-	-	-	-
BEN 0374	Gallus gallus	Spain	APEC	B2 IX	18	+	+	+	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-
BEN 0079	Gallus gallus	France	APEC	B2 IX	18	+	+	+	+	+	_	_	_	-	+	+	+	+	-	-	_	_	-	_	-	_	-	_	-

Colors correspond to the different PAIs and plasmid: orange, PAI II_{J96}; green, PAI III₅₃₆; violet, HPI; grey, plasmid origin.

_

2 Legend of Figures

3

Fig. 1. Factorial analysis of correspondence (FAC) of the 234 *E. coli* strains. Projections of
the 35 variables: human/animal origin, commensal, ExPEC and InPEC characters, 7
phylogenetic groups (A, B1, B2, C, D, E and F) and ungrouped strains (UG) and 23 VFs (see
materials and methods) projected onto the F1/F2 plane calculated in the FAC (22). AnAd:
Animal adhesin.

9

Fig. 2. Factorial analysis of correspondence (FAC) of the 234 *E. coli* strains. Projections of (A) the ExPEC strains (black symbols) of animal (triangle) and human (square) origin and of the non ExPEC strains (white symbols) and (B) the InPEC strains of animal (triangle) and human (square) origin and of the non InPEC strains (white symbols) on the F1/F2 plane calculated in the FAC (22).

15

Fig. 3. Phylogenetic tree of the 234 *E. coli* strains studied, reconstructed from the partial sequences of 8 housekeeping genes (<u>www.pasteur.fr/mlst</u>) by PHYML (23) and rooted on *E. fergusonii*. Bootstrap values are indicated at the corresponding nodes only when they exceed 70%. Pathogenic strains are indicated in red, whereas commensal strains are shown in black. The name of the strain is followed by H (for human origin) or A (for animal origin). The colors of the terminal branches of the tree are as follows: orange (F group), red (B2 group), violet (E group), light blue (C group), blue (A group) and green (B1 group).

1 Supplementary online material

- 2
- 3 Table S1. List and principal characteristics of the *E. coli* strains studied.
- 4 Table S2. List of the oligonucleotides used for the detection of the intraintestinal virulence
- 5 factors.
- 6 Table S3. List of the oligonucleotides used in the O-typing PCR assay.
- 7

1 References

- 2
- 3
- 4 1. Achtman, M., M. Heuzenroeder, B. Kusecek, H. Ochman, D. Caugant, R. K. 5 Selander, V. Vaisanen-Rhen, T. K. Korhonen, S. Stuart, F. Orskov, and et al. 6 1986. Clonal analysis of Escherichia coli O2:K1 isolated from diseased humans and 7 animals. Infect Immun 51:268-76. 8 2. Bertin, Y., C. Martin, E. Oswald, and J. P. Girardeau. 1996. Rapid and specific 9 detection of F17-related pilin and adhesin genes in diarrheic and septicemic 10 Escherichia coli strains by multiplex PCR. J Clin Microbiol 34:2921-8. Bingen, E., B. Picard, N. Brahimi, S. Mathy, P. Desjardins, J. Elion, and E. 11 3. 12 Denamur. 1998. Phylogenetic analysis of Escherichia coli strains causing neonatal 13 meningitis suggests horizontal gene transfer from a predominant pool of highly 14 virulent B2 group strains. J Infect Dis 177:642-50. 15 4. Boerlin, P., R. Travis, C. L. Gyles, R. Reid-Smith, N. Janecko, H. Lim, V. Nicholson, S. A. McEwen, R. Friendship, and M. Archambault. 2005. 16 Antimicrobial resistance and virulence genes of Escherichia coli isolates from swine 17 18 in Ontario. Appl Environ Microbiol 71:6753-61. 19 5. Chen, S. L., C. S. Hung, J. Xu, C. S. Reigstad, V. Magrini, A. Sabo, D. Blasiar, T. 20 Bieri, R. R. Meyer, P. Ozersky, J. R. Armstrong, R. S. Fulton, J. P. Latreille, J. Spieth, T. M. Hooton, E. R. Mardis, S. J. Hultgren, and J. I. Gordon. 2006. 21 22 Identification of genes subject to positive selection in uropathogenic strains of 23 Escherichia coli: a comparative genomics approach. Proc Natl Acad Sci U S A 24 103:5977-82. 25 6. Cherifi, A., M. Contrepois, B. Picard, P. Goullet, I. Orskov, and F. Orskov. 1994. 26 Clonal relationships among Escherichia coli serogroup O78 isolates from human and 27 animal infections. J Clin Microbiol 32:1197-202. Cherifi, A., M. Contrepois, B. Picard, P. Goullet, I. Orskov, F. Orskov, and J. De 28 7. 29 Rycke. 1991. Clonal relationships among Escherichia coli serogroup 06 isolates from 30 human and animal infections. FEMS Microbiol Lett 64:225-30. 31 8. Clawson, M. L., J. E. Keen, T. P. Smith, L. M. Durso, T. G. McDaneld, R. E. 32 Mandrell, M. A. Davis, and J. L. Bono. 2009. Phylogenetic classification of 33 Escherichia coli O157:H7 strains of human and bovine origin using a novel set of 34 nucleotide polymorphisms. Genome Biol 10:R56. 35 9. Clermont, O., S. Bonacorsi, and E. Bingen. 2000. Rapid and simple determination 36 of the Escherichia coli phylogenetic group. Appl Environ Microbiol 66:4555-8. 37 Clermont, O., J. R. Johnson, M. Menard, and E. Denamur. 2007. Determination of 10. 38 Escherichia coli O types by allele-specific polymerase chain reaction: application to 39 the O types involved in human septicemia. Diagn Microbiol Infect Dis 57:129-36. Clermont, O., M. Lescat, C. L. O'Brien, D. M. Gordon, O. Tenaillon, and E. 40 11. 41 Denamur. 2008. Evidence for a human-specific Escherichia coli clone. Environ 42 Microbiol 10:1000-6. 43 12. Diard, M., S. Baeriswyl, O. Clermont, S. Gouriou, B. Picard, F. Taddei, E. 44 Denamur, and I. Matic. 2007. Caenorhabditis elegans as a simple model to study 45 phenotypic and genetic virulence determinants of extraintestinal pathogenic Escherichia coli. Microbes Infect 9:214-23. 46

1	13.	Dow, M. A., I. Toth, P. Alexa, M. Davies, A. Malik, E. Oswald, and B. Nagy.
2		2005. Predominance of afr2 and ral fimbrial genes related to those encoding the K88
3		and CS31A fimbrial adhesins in enteropathogenic Escherichia coli isolates from
4		rabbits with postweaning diarrhea in Central Europe. J Clin Microbiol 43:1366-71.
5	14.	Escobar-Paramo, P., O. Clermont, A. B. Blanc-Potard, H. Bui, C. Le Bouguenec,
6		and E. Denamur. 2004. A specific genetic background is required for acquisition and
7		expression of virulence factors in Escherichia coli. Mol Biol Evol 21 :1085-94.
8	15.	Escobar-Paramo, P., A. Le Menac'h, T. Le Gall, C. Amorin, S. Gouriou, B.
9		Picard, D. Skurnik, and E. Denamur. 2006. Identification of forces shaping the
10		commensal Escherichia coli genetic structure by comparing animal and human
11		isolates. Environ Microbiol 8:1975-84.
12	16.	Ewers, C., G. Li, H. Wilking, S. Kiessling, K. Alt, E. M. Antao, C. Laturnus, I.
13		Diehl, S. Glodde, T. Homeier, U. Bohnke, H. Steinruck, H. C. Philipp, and L. H.
14		Wieler. 2007. Avian pathogenic, uropathogenic, and newborn meningitis-causing
15		Escherichia coli: how closely related are they? Int J Med Microbiol 297: 163-76.
16	17.	Feng, P. C., S. R. Monday, D. W. Lacher, L. Allison, A. Siitonen, C. Keys, M.
17		Eklund, H. Nagano, H. Karch, J. Keen, and T. S. Whittam. 2007. Genetic diversity
18		among clonal lineages within Escherichia coli O157:H7 stepwise evolutionary model.
19		Emerg Infect Dis 13:1701-6.
20	18.	Franck, S. M., B. T. Bosworth, and H. W. Moon. 1998. Multiplex PCR for
21		enterotoxigenic, attaching and effacing, and Shiga toxin-producing Escherichia coli
22		strains from calves. J Clin Microbiol 36: 1795-7.
23	19.	Girardeau, J. P., A. Dalmasso, Y. Bertin, C. Ducrot, S. Bord, V. Livrelli, C.
24		Vernozy-Rozand, and C. Martin. 2005. Association of virulence genotype with
25		phylogenetic background in comparison to different seropathotypes of Shiga toxin-
26		producing Escherichia coli isolates. J Clin Microbiol 43: 6098-107.
27	20.	Girardeau, J. P., L. Lalioui, A. M. Said, C. De Champs, and C. Le Bouguenec.
28		2003. Extended virulence genotype of pathogenic Escherichia coli isolates carrying
29		the afa-8 operon: evidence of similarities between isolates from humans and animals
30		with extraintestinal infections. J Clin Microbiol 41:218-26.
31	21.	Gordon, D. M., O. Clermont, H. Tolley, and E. Denamur. 2008. Assigning
32		Escherichia coli strains to phylogenetic groups: multi-locus sequence typing versus
33		the PCR triplex method. Environ Microbiol 10: 2484-96.
34	22.	Greenacre, M. 1992. Correspondence analysis in medical research. Stat Methods
35		Med Res 1: 97-117.
36	23.	Guindon, S., F. Lethiec, P. Duroux, and O. Gascuel. 2005. PHYML Onlinea web
37		server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res
38		33: W557-9.
39	24.	Imberechts, H., N. Van Pelt, H. De Greve, and P. Lintermans. 1994. Sequences
40		related to the major subunit gene fedA of F107 fimbriae in porcine Escherichia coli
41		strains that express adhesive fimbriae. FEMS Microbiol Lett 119: 309-14.
42	25.	Jaureguy, F., L. Landreau, V. Passet, L. Diancourt, E. Frapy, G. Guigon, E.
43		Carbonnelle, O. Lortholary, O. Clermont, E. Denamur, B. Picard, X. Nassif, and
44		S. Brisse. 2008. Phylogenetic and genomic diversity of human bacteremic Escherichia
45		coli strains. BMC Genomics 9:560.
46	26.	Johnson, J. R., O. Clermont, M. Menard, M. A. Kuskowski, B. Picard, and E.
47		Denamur. 2006. Experimental mouse lethality of Escherichia coli isolates, in relation
48		to accessory traits, phylogenetic group, and ecological source. J Infect Dis 194:1141-
49		50.

1	27.	Johnson, J. R., P. Delavari, A. L. Stell, T. S. Whittam, U. Carlino, and T. A.
2		Russo. 2001. Molecular comparison of extraintestinal Escherichia coli isolates of the
3		same electrophoretic lineages from humans and domestic animals. J Infect Dis
4		183: 154-9.
5	28.	Johnson, T. J., S. Kariyawasam, Y. Wannemuehler, P. Mangiamele, S. J.
6		Johnson, C. Doetkott, J. A. Skyberg, A. M. Lynne, J. R. Johnson, and L. K.
7		Nolan. 2007. The genome sequence of avian pathogenic Escherichia coli strain
8		O1:K1:H7 shares strong similarities with human extraintestinal pathogenic E. coli
9		genomes. J Bacteriol 189:3228-36.
10	29.	Johnson, T. J., Y. Wannemuehler, S. J. Johnson, A. L. Stell, C. Doetkott, J. R.
11		Johnson, K. S. Kim, L. Spanjaard, and L. K. Nolan. 2008. Comparison of
12		extraintestinal pathogenic Escherichia coli strains from human and avian sources
13		reveals a mixed subset representing potential zoonotic pathogens. Appl Environ
14		Microbiol 74: 7043-50.
15	30.	Kaper, J. B., J. P. Nataro, and H. L. Mobley. 2004. Pathogenic Escherichia coli.
16		Nat Rev Microbiol 2: 123-40.
17	31.	Lawrence, J. G., H. Ochman, and D. L. Hartl. 1991. Molecular and evolutionary
18		relationships among enteric bacteria. J Gen Microbiol 137:1911-21.
19	32.	Le Gall, T., O. Clermont, S. Gouriou, B. Picard, X. Nassif, E. Denamur, and O.
20		Tenaillon. 2007. Extraintestinal virulence is a coincidental by-product of
21		commensalism in B2 phylogenetic group Escherichia coli strains. Mol Biol Evol
22		24: 2373-84.
23	33.	MacLean, D., J. D. Jones, and D. J. Studholme. 2009. Application of 'next-
24		generation' sequencing technologies to microbial genetics. Nat Rev Microbiol 7:287-
25		96.
26	34.	Mariani-Kurkdjian, P., E. Denamur, A. Milon, B. Picard, H. Cave, N. Lambert-
27		Zechovsky, C. Loirat, P. Goullet, P. J. Sansonetti, and J. Elion, 1993.
28		Identification of a clone of Escherichia coli O103:H2 as a potential agent of
29		hemolytic-uremic syndrome in France. J Clin Microbiol 31 :296-301.
30	35.	Maynard, C., S. Bekal, F. Sanschagrin, R. C. Levesque, R. Brousseau, L. Masson,
31		S. Lariviere, and J. Harel. 2004. Heterogeneity among virulence and antimicrobial
32		resistance gene profiles of extraintestinal Escherichia coli isolates of animal and
33		human origin. J Clin Microbiol 42: 5444-52.
34	36.	Moissenet, D., B. Salauze, O. Clermont, E. Bingen, G. Arlet, E. Denamur, A.
35		Merens, D. Mitanchez, and H. Vu-Thien. Meningitis caused by Escherichia coli
36		producing TEM-52 extended-spectrum beta-lactamase within an extensive outbreak in
37		a neonatal ward: epidemiological investigation and characterization of the strain. J
38		Clin Microbiol 48: 2459-63.
39	37.	Mora, A., C. Lopez, G. Dabhi, M. Blanco, J. E. Blanco, M. P. Alonso, A. Herrera,
40		R. Mamani, S. Bonacorsi, M. Moulin-Schouleur, and J. Blanco. 2009.
41		Extraintestinal pathogenic Escherichia coli O1:K1:H7/NM from human and avian
42		origin: detection of clonal groups B2 ST95 and D ST59 with different host
43		distribution. BMC Microbiol 9: 132.
44	38.	Moulin-Schouleur, M., M. Reperant, S. Laurent, A. Bree, S. Mignon-Grasteau, P.
45	20.	Germon, D. Rasschaert, and C. Schouler. 2007. Extraintestinal pathogenic
46		Escherichia coli strains of avian and human origin: link between phylogenetic
47		relationships and common virulence patterns. I Clin Microbiol 45:3366-76
48	39.	Moulin-Schouleur, M., C. Schouler, P. Tailliez, M. R. Kao, A. Bree, P. Germon.
49		E. Oswald, J. Mainil, M. Blanco, and J. Blanco . 2006. Common virulence factors
-		

1 and genetic relationships between O18:K1:H7 Escherichia coli isolates of human and 2 avian origin. J Clin Microbiol 44:3484-92. 3 40. Newton, H. J., J. Sloan, D. M. Bulach, T. Seemann, C. C. Allison, M. Tauschek, 4 R. M. Robins-Browne, J. C. Paton, T. S. Whittam, A. W. Paton, and E. L. 5 Hartland. 2009. Shiga toxin-producing Escherichia coli strains negative for locus of 6 enterocyte effacement. Emerg Infect Dis 15:372-80. 7 41. Ochman, H., and R. K. Selander. 1984. Standard reference strains of Escherichia 8 coli from natural populations. J Bacteriol 157:690-3. 9 42. Ogura, Y., T. Ooka, A. Iguchi, H. Toh, M. Asadulghani, K. Oshima, T. Kodama, 10 H. Abe, K. Nakayama, K. Kurokawa, T. Tobe, M. Hattori, and T. Hayashi. 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and 11 12 non-O157 enterohemorrhagic Escherichia coli. Proc Natl Acad Sci U S A 106:17939-13 44. 14 Picard, B., J. S. Garcia, S. Gouriou, P. Duriez, N. Brahimi, E. Bingen, J. Elion, 43. 15 and E. Denamur. 1999. The link between phylogeny and virulence in Escherichia coli extraintestinal infection. Infect Immun 67:546-53. 16 17 Pradel, N., K. Boukhors, Y. Bertin, C. Forestier, C. Martin, and V. Livrelli. 2001. 44. 18 Heterogeneity of Shiga toxin-producing Escherichia coli strains isolated from 19 hemolytic-uremic syndrome patients, cattle, and food samples in central France. Appl 20 Environ Microbiol 67:2460-8. 21 45. Rasko, D. A., M. J. Rosovitz, G. S. Myers, E. F. Mongodin, W. F. Fricke, P. 22 Gajer, J. Crabtree, M. Sebaihia, N. R. Thomson, R. Chaudhuri, I. R. Henderson, 23 V. Sperandio, and J. Ravel. 2008. The pangenome structure of Escherichia coli: 24 comparative genomic analysis of E. coli commensal and pathogenic isolates. J 25 Bacteriol 190:6881-93. 26 Reid, S. D., C. J. Herbelin, A. C. Bumbaugh, R. K. Selander, and T. S. Whittam. 46. 2000. Parallel evolution of virulence in pathogenic Escherichia coli. Nature 406:64-7. 27 28 47. Rhen, M., P. Klemm, and T. K. Korhonen. 1986. Identification of two new 29 hemagglutinins of Escherichia coli, N-acetyl-D-glucosamine-specific fimbriae and a 30 blood group M-specific agglutinin, by cloning the corresponding genes in Escherichia 31 coli K-12. J Bacteriol 168:1234-42. 32 Ron, E. Z. 2006. Host specificity of septicemic Escherichia coli: human and avian 48. 33 pathogens. Curr Opin Microbiol 9:28-32. 34 49. Savageau, M. A. 1983. Escherichia coli habitats, cell types, and moleculat 35 mechanisms of gene control. Am Nat 122:732-744. Tenaillon, O., D. Skurnik, B. Picard, and E. Denamur. 2010. The population 36 50. 37 genetics of commensal Escherichia coli. Nat Rev Microbiol 8:207-17. 38 Touchon, M., C. Hoede, O. Tenaillon, V. Barbe, S. Baeriswyl, P. Bidet, E. Bingen, 51. 39 S. Bonacorsi, C. Bouchier, O. Bouvet, A. Calteau, H. Chiapello, O. Clermont, S. 40 Cruveiller, A. Danchin, M. Diard, C. Dossat, M. E. Karoui, E. Frapy, L. Garry, J. M. Ghigo, A. M. Gilles, J. Johnson, C. Le Bouguenec, M. Lescat, S. Mangenot, V. 41 42 Martinez-Jehanne, I. Matic, X. Nassif, S. Oztas, M. A. Petit, C. Pichon, Z. Rouy, 43 C. S. Ruf, D. Schneider, J. Tourret, B. Vacherie, D. Vallenet, C. Medigue, E. P. 44 Rocha, and E. Denamur. 2009. Organised genome dynamics in the Escherichia coli 45 species results in highly diverse adaptive paths. PLoS Genet 5:e1000344. 46 52. Wirth, T., D. Falush, R. Lan, F. Colles, P. Mensa, L. H. Wieler, H. Karch, P. R. 47 Reeves, M. C. Maiden, H. Ochman, and M. Achtman. 2006. Sex and virulence in 48 Escherichia coli: an evolutionary perspective. Mol Microbiol 60:1136-51.

1	53.	Wu, G., B. Carter, M. Mafura, E. Liebana, M. J. Woodward, and M. F. Anjum.
2		2008. Genetic diversity among Escherichia coli O157:H7 isolates and identification of
3		genes linked to human infections. Infect Immun 76:845-56.
4	54.	Xie, Y., V. Kolisnychenko, M. Paul-Satyaseela, S. Elliott, G. Parthasarathy, Y.
5		Yao, G. Plunkett, 3rd, F. R. Blattner, and K. S. Kim. 2006. Identification and
6		characterization of Escherichia coli RS218-derived islands in the pathogenesis of E.
7		coli meningitis. J Infect Dis 194: 358-64.
8	55.	Zhao, L., S. Gao, H. Huan, X. Xu, X. Zhu, W. Yang, Q. Gao, and X. Liu. 2009.
9		Comparison of virulence factors and expression of specific genes between
10		uropathogenic Escherichia coli and avian pathogenic E. coli in a murine urinary tract
11		infection model and a chicken challenge model. Microbiology 155 :1634-44.
12		
13		







٩,

 \sim



Factor 2



Fig. 2

