

Research article

Open Access

Two distinct groups of porcine enteropathogenic *Escherichia coli* strains of serogroup O45 are revealed by comparative genomic hybridization and virulence gene microarray

Guillaume Bruant^{†1}, Yongxiang Zhang^{†2}, Philippe Garneau¹, Justin Wong², Chad Laing², John M Fairbrother¹, Victor PJ Gannon² and Josée Harel^{*1}

Address: ¹Groupe de Recherche sur les Maladies Infectieuses du Porc, Faculté de médecine vétérinaire, Université de Montréal, 3200 rue Sicotte, Saint-Hyacinthe, Québec J2S 7C6, Canada and ²Laboratory for Foodborne Zoonoses, Public Health Agency of Canada, Lethbridge, Alberta, T1J 3Z4, Canada

Email: Guillaume Bruant - guillaume.bruant@nrc-cnrc.gc.ca; Yongxiang Zhang - Yongxiang.Zhang@inspection.gc.ca; Philippe Garneau - philippe.garneau@umontreal.ca; Justin Wong - justin@brc.ubc.ca; Chad Laing - Chad.Laing@inspection.gc.ca; John M Fairbrother - john.morris.fairbrother@umontreal.ca; Victor PJ Gannon - Vic.Gannon@inspection.gc.ca; Josée Harel* - josee.harel@umontreal.ca

* Corresponding author †Equal contributors

Published: 26 August 2009

Received: 23 April 2009

BMC Genomics 2009, 10:402 doi:10.1186/1471-2164-10-402

Accepted: 26 August 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/402>

© 2009 Bruant et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Porcine enteropathogenic *Escherichia coli* (PEPEC) strains of serogroup O45 cause post-weaning diarrhea and produce characteristic attaching and effacing (A/E) lesions. Most O45 PEPEC strains possess the locus of enterocyte effacement (LEE), encoding the virulence factors required for production of A/E lesions, and often possess the *paa* gene, which is thought to contribute to the early stages of PEPEC pathogenicity. In this study, nine O45 PEPEC strains and a rabbit enteropathogenic (REPEC) strain, known to produce A/E lesions *in vivo*, were characterized using an *E. coli* O157-*E. coli* K12 whole genome microarray and a virulence gene-specific microarray, and by PCR experiments.

Results: Based on their virulence gene profiles, the 10 strains were considered to be atypical EPEC. The differences in their genomes pointed to the identification of two distinct evolutionary groups of O45 PEPEC, Groups I and II, and provided evidence for a contribution of these genetic differences to their virulence in pigs. Group I included the REPEC strain and four O45 PEPEC strains known to induce severe A/E lesions in challenged pigs whereas Group II was composed of the five other O45 PEPEC strains, which induced less severe or no A/E lesions in challenged pigs. Significant differences between Groups I and II were found with respect to the presence or absence of 50 O-islands (OIs) or S-loops and 13 K-islands (KIs) or K-loops, including the virulence-associated islands OI#1 (S-loop#1), OI#47 (S-loop#71), OI#57 (S-loop#85), OI#71 (S-loop#108), OI#115, OI#122, and OI#154 (S-loop#253).

Conclusion: We have genetically characterized a collection of O45 PEPEC strains and classified them into two distinct groups. The differences in their virulence gene and genomic island content may influence the pathogenicity of O45 PEPEC strains, and explain why Group I O45 PEPEC strains induced more severe A/E lesions in explants and challenged pigs than Group II strains.

Background

Escherichia coli of serogroup O45 may be isolated both in intestinal and extraintestinal sites, although they have been only sporadically described in the latter [1-3]. On the other hand, intestinal *E. coli* strains have been more frequently identified as belonging to this serogroup. Intestinal O45 *E. coli* strains have been isolated from animals and humans and have been classified as both enterotoxigenic (ETEC) and attaching and effacing (AEEC) *E. coli*, the latter including both enterohemorrhagic (EHEC) and enteropathogenic (EPEC) *E. coli* [4-6]. Serogroup O45 is particularly important among porcine EPEC (PEPEC) strains which cause post-weaning diarrhea (PWD) characterized by specific attaching and effacing (A/E) lesions [7-9]). Most O45 PEPEC strains possess the locus of enterocyte effacement (LEE) pathogenicity island, which contains virulence genes necessary for the production of A/E lesions. They also often possess the *paa* gene (for porcine A/E associated gene), which encodes a virulence factor involved in the A/E phenotype and is thought to contribute to the early stages of PEPEC pathogenicity [10]. These strains also have the ability to produce A/E lesions in experimentally inoculated newborn gnotobiotic piglets and in a homologous *in vitro* model using newborn piglet ileal explants, as well as to adhere to PK15 porcine kidney cells *in vitro* [10-14].

Genomic islands (GIs) such as LEE are regions of bacterial genomes that have been acquired by horizontal gene transfer and often contain blocks of genes that function together in specific processes. When the genomes of the two *E. coli* O157:H7 strains EDL933 and Sakai were compared with that of *E. coli* K12 strain MG1655, the GIs found to be present in strains EDL933 and Sakai but absent in strain MG1655 were named O-islands (OIs) and Sakai loops (S-loops), respectively. The GIs found to be present in *E. coli* K12 but absent from the two *E. coli* O157:H7 strains were named K-islands (KIs) and K-loops, respectively [15-17]. GIs related to the virulence of a pathogen are also referred to as pathogenicity islands (PAIs) [18]. In *E. coli* O157:H7 strain EDL933, several large OIs encode virulence or putative virulence factors. These OIs include OI#45 (S-loop#69) and OI#93 (S-loop#153) for Shiga toxin 2 and 1, respectively, OI#148 (S-loop#244) for LEE, and OI#57 (S-loop#85) for *paa* [16,17].

A recent microarray-based study has catalogued genomic alterations in a collection of *E. coli* O157:H7 strains, particularly in GIs, suggesting the existence of two dominant lineages, with characteristics that are unique to each of them [19]. Previous studies performed on various AEEC strains have also shown that, depending on their pathotype and host specificity, strains can show variations in their LEE sequences as well as in the site of insertion of LEE in the chromosome [14,20,21]. The purpose of the

present study was to examine the genotypic differences, particularly in LEE sequences and chromosomal insertion sites, and in the presence or absence of non LEE-encoded virulence factors, such as *Paa*, among a collection of O45 PEPEC strains which have been previously shown to induce A/E lesions in pigs. In this study, we have characterized O45 PEPEC strains using a DNA-microarray designed specifically for detection of *E. coli* virulence genes [22] and compared their genomes using comparative genomic hybridization (CGH) and PCR. We identified two distinct groups of PEPEC O45 strains, between which there were significant variations in GI content.

Methods

Bacterial strains and preparation of genomic DNA

Nine O45 PEPEC strains, which were isolated at the Faculté de médecine vétérinaire, Saint-Hyacinthe, Québec, Canada, from pigs with PWD [13] were used for the microarray studies (Table 1). These strains were selected based, i) on their ability to produce or not A/E lesions in challenged pigs [13] and in an homologous *ex vivo* model using newborn piglet ileal explants (data not shown), ii) and on the severity of the A/E manifestation they produced [13] (Table 1). Because of its genetic and phenotypic similarities with the O45 PEPEC strains [23], the O103 rabbit EPEC (REPEC) strain E22, provided by Eric Oswald (INRA, Toulouse, France) [24,25], was included in the study. Five *E. coli* reference strains were used as controls in PCR experiments: the two O157:H7 *E. coli* strains EDL933 and Sakai, the K12 strain MG1655, the uropathogenic (UPEC) strain CFT073 and the REPEC strain RDEC-1.

For DNA preparations, strains were grown overnight in 45 mL of Brain-Heart-Infusion (BHI) broth at 37°C. The cultures were centrifuged at 8000 rpm for 10 minutes and the pellet was dissolved in 15 mL of 10 mM NaCl, 20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 100 µg/mL proteinase K and 0.5% SDS. This suspension was incubated at 50°C for 2 h and DNA was extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1). Following centrifugation for 10 min at 8000 rpm, the upper phase was removed and precipitated by adding 0.1 volume of 3 M NaOAc (pH 5.2) and 2 volumes of 99% ethanol. The DNA precipitate was then spooled out of the solution using a sterile glass rod, washed with 70% ethanol, and dissolved in 5 mL of TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) buffer.

E. coli DNA microarrays

For the whole genome microarray (named *E. coli* O157:H7 microarray), Corning Ultra-Gap II slides (Corning, Acton, MA) were spotted with the MWG *E. coli* O157:H7 array set (MWG Biotech). The MWG array set consists of 6167 50-mer oligonucleotides covering the

Table 1: Characteristics of O45 PEPEC strains and REPEC strain E22 used in this study.

Name ^a	Serogroup	Origin	A/E lesions in explants ^b	A/E lesions in challenged piglets ^c		
				Extent ^d	Site ^e	Inflammation ^f
E22 ^g	O103:H2	Rabbit	ND	ND	ND	ND
ECL1001 (86-1390)	O45:KE65	Pig	+	++++	I/Ce/C-P/C-D	S
ECL2004 (81-4420)	O45:KE65	Pig	+	++++	I/Ce/C-P/C-D	L
ECL2017 (86-4220)	O45:KE65	Pig	+	+++	I/Ce/C-P/C-D	M
ECL2033 (91-19-172)	O45:KE65	Pig	+	++++	I/Ce/C-P/C-D	-
ECL2019 (88-1861)	O45:KE65	Pig	-	-	-	-
ECL2020 (88-4299)	O45:KE65	Pig	+	++	C-P/C-D	L
ECL2027 (89-56-196)	O45:KE65	Pig	-	-	-	-
ECL2076 ^h (87-4725)	O45:KE65	Pig	+	-	-	-
ECL2078 (83-2315)	O45:KE65	Pig	+	++	D//I/Ce/C-P/C-D	L

^a Former names used in a previous study by Zhu *et al.* [13] are indicated in parentheses.

^b Unpublished data from studies performed in our laboratory. +, strain tested produced A/E lesions in explants; -, no lesion observed; ND, not determined.

^c Data from a previous study by Zhu *et al.* [13].

^d +++++, extensive bacterial colonization and severe effacement of microvilli; +++, large areas of bacterial colonization and heavy effacement; ++, focal lesions; +, small scattered focal lesions; -, no lesions observed; ND, not determined.

^e A/E lesions were observed in duodenum (D), ileum (I), cecum (Ce), proximal colon (C-P), and distal colon (C-D); ND, not determined.

^f L, light inflammatory response; M, mild inflammatory response; S, severe inflammatory response; -, no inflammatory response; ND, not determined.

^g REPEC reference strain [24,25].

^h Strain ECL2076 produced A/E lesions in explants but was not included in the study by Zhu *et al.* [13].

whole genomes of *E. coli* K-12 strain MG1655 [15] and *E. coli* O157:H7 strains Sakai (RIMD 0509952) [16] and EDL933 (ATCC700927) [17].

The *E. coli* virulence microarray used in this study was derived from the one previously developed by Bruant *et al.* and included 315 70-mer oligonucleotides specific for 189 *E. coli* virulence or putative virulence genes or markers found in various intestinal and extraintestinal *E. coli* strains of all known pathotypes [22]. Probes were specific for genes encoding adhesins; toxins; bacteriocins; anti-aggregative factors; autotransporters; capsular, flagellar, and somatic antigens; hemolysins; invasins; iron acquisition systems or transport proteins; and outer membrane proteins, as well as other genes recently shown to be associated with virulence in *E. coli*. This microarray also detected genetic variants of particular genes, such as the intimin-encoding gene *eae* (variants alpha, alpha2, beta, beta2, delta, epsilon, epsilon2, eta, gamma, gamma2, iota, iota2, lambda, mu, nu, pi, xi, and zeta), *espA* (variants *espA1*, *espA2*, and *espA3*), *espB* (variants *espB1*, *espB2*, and *espB3*), and *tir* (variants *tir-1*, *tir-2*, and *tir-3*) from the LEE. Oligonucleotides specific for three variants of the major fimbrial subunit of the long polar fimbria (LPF)-encoding gene *lpfA* were also included. These were based on sequences from the *lpfA* genes of EPEC strains of serogroup O113 (*lpfA*_{O113}), OI#141 from *E. coli* of serotype O157:H7 (*lpfA1*), and REPEC strains and *E. coli* of serogroup O26 (*lpfA*_{R141}).

Microarray hybridizations

Prior to *E. coli* O157:H7 microarray hybridization, each array was pre-hybridized at 50°C in a solution of 5 × SSC, 0.1% SDS and 0.1% BSA for 1 h. Following this step, arrays were washed completely in dH₂O, rinsed with isopropanol, and then centrifuged and dried. For hybridization, 5 µg of test genomic DNA were digested with *EcoRV* and *PstI* restriction enzymes, 3 µg of which were labeled with ULYSIS Alexa Fluor 647 dye (Invitrogen, Burlington, ON). Genomic DNA from strains MG1655, Sakai and EDL933 was digested in an analogous fashion, and 1 µg of the preparation from each strain was combined and labeled with Alexa Fluor 546 dye (Invitrogen, Burlington, ON). This labeled genomic DNA mixture was then used as a reference for all hybridizations. Unincorporated dye was removed using Qiaquick PCR purification kits (Qiagen, Mississauga, ON), according to the manufacturer's instructions, and DNA was eluted in 30 µl of 0.1 × TE buffer. Labeled DNA was vacuum-dried and resuspended in 20 µl of dH₂O. A 70 µl hybridization solution consisting of 30% formamide, 5 × SSC, 0.1% SDS, 0.1 mg/ml sonicated Salmon sperm DNA, and equal amounts of test and reference labeled DNAs, each containing at least 30 pmol of incorporated dye, was denatured at 95°C for 5 min and briefly centrifuged to collect all the contents. DNA preparations were then hybridized overnight (16 h) at 42°C. After hybridization, arrays were washed according to the modified Corning method (Corning). Arrays were then scanned with a GenePix 4000B scanner (Axon

Instruments, Redwood City, CA) and processed using GenePix Pro 5.0. Two slides were hybridized per strain with a dye-swap repeat per slide.

Hybridizations on *E. coli* virulence microarrays were performed as described previously [22]. Arrays were scanned with a ScanArray® Lite fluorescent microarray analysis system (Canberra-Packard Canada, Montreal, Quebec) and acquisition and quantification of fluorescent spot intensities were performed using the ScanArray Express® software version 2.1 (Perkin-Elmer, Foster City, CA, USA).

Microarray data analysis

Data obtained from *E. coli* O157:H7 microarrays were normalized using the Ratio-based and Lowess methods in Acuity 3.1 (Axon instruments) before analysis. The normalized data for all strains were converted into \log_2 (Fluor 647/Fluor 546) in Acuity 3.1 and subsequently analyzed in Microsoft Excel. Control, blank, and test spots with a mean intensity below that of the mean of all negative controls were removed from the analysis. The arithmetic mean of the remaining spots across the four duplicates was calculated to construct the dataset. GACK (for Genomotyping Analysis by Charles Kim) [26], was used to generate a cut-off value determining the presence or absence of genes, and a dendrogram using the Euclidean distance metric with average linkage was created with tMEV v4.1 [27].

For the data obtained from *E. coli* virulence microarrays, the local background was subtracted from the recorded spot intensities. The median value of each set of triplicate spotted oligonucleotides was then compared to the median value of the negative control spots present on the array. Oligonucleotides with a signal-to-noise fluorescence ratio greater than 2.0 were considered as positive.

Microarray data accession number

The microarray data have been deposited in NCBI's Gene Expression Omnibus (GEO accession number GSE17036) <http://www.ncbi.nlm.nih.gov/projects/geo/>.

PCR experiments

PCR experiments were performed for the nine O45 PEPEC strains and the REPEC strain E22 to determine the localization in their chromosome of the LEE and of the OI#122, as well as the integrity of the OI#122 and of the secondary type III secretion system gene cluster designated ETT2 (for *E. coli* type III secretion 2). All PCR experiments were performed as described in previous studies carried out on the LEE, OI#122 and ETT2 gene clusters (Additional file 1: Table S1 [21,23,28-31]).

PCR experiments were also performed for *nleA* and *nleC* genes. The pairs of primers used were *nleA*-F (ACCG-

CAATCCGAATTACCTC) – *nleA*-R (TCCATTGCGCGTATATCAGC) and ECs1812F (CTGTCCAACAGGGATAC) – ECs1812R (CCGCAATCCGAATTACC) for *nleA*, and *nleC*-F (AAGTGTAATACGCGCCGTCC) – *nleC*-R (ATCAGGACTCGCCTCATATC) and ECs0847F (CCCATTGCTCCTAATCG) – ECs0847R (CAGCGGAATACTCTGTG) for *nleC*. The conditions for amplifications were an initial denaturation of 95°C for 5 min, followed by 30 cycles of 95°C for 30 s; 55°C for 30 s; 72°C for 80 s and a final elongation of 72°C for 10 min.

Results

Characterization of O45 PEPEC strains using the *E. coli* virulence microarray

All O45 PEPEC strains and the REPEC strain E22 were characterized using the *E. coli* virulence microarray described previously, which includes probes targeting virulence genes generally found in AEEC but also virulence genes from the other *E. coli* virotypes [22]. All strains possessed their own specific virulence gene profile but were all classified as atypical EPEC (Additional file 2: Table S2 [32]). They all possessed the LEE genes and shared the same LEE profile: *eae*(β) – *espA* group I – *espB* group III – *tir* group I. In addition, each strain lacked the Shiga toxin 2 encoding genes *stx2A* and *stx2B*, as well as the bundle forming pili (BFP) encoding gene *bfpA* and the *E. coli* adherence factor (EAF) virulence plasmid marker *eaf*. Remarkably, all O45 PEPEC strains and REPEC strain E22, although *stx1A*-negative, gave a positive hybridization for the *stx1B* gene, which encodes the B subunit of EHEC Shiga-like toxin 1 and which is generally absent in EPEC strains. However, the presence of the *stx1B* gene was not confirmed in PCR experiments (data not shown).

The *E. coli* strains could be classified into two distinct groups according to their virulence gene pattern (Additional file 2: Table S2 [32]). Group I included the four PEPEC strains ECL1001, ECL2004, ECL2017, and ECL2033, and REPEC strain E22. Group II included the five other PEPEC strains ECL2019, ECL2020, ECL2027, ECL2076, and ECL2078. Results obtained with the *E. coli* virulence microarray identified 19 virulence genes that showed a non-random distribution between Group I and Group II strains (Additional file 2: Table S2 [32]). Genes *b1121* (encoding a hypothetical protein YcfZ), *set* (encoding a probable enterotoxin, also named *ent*), *tspE4.C2* (an anonymous fragment), *efa1* (encoding the EHEC factor for adherence Efa1), and *paa* were present in all Group I strains, including REPEC strain E22, but absent from all Group II strains. The temperature sensitive hemagglutinin encoding gene *tsh*, the yersiniabactin-related genes *fyuA*, *irp1* and *irp2*, as well as the PAI-associated gene *malX* were present in all O45 PEPEC strains from Group I but in neither REPEC strain E22 nor Group II strains. The heat-stable enterotoxin encoding gene *astA* was present in all O45

PEPEC Group I strains and in Group II strain ECL2020 but absent from all other Group II strains and REPEC strain E22.

The genes *aidal* (encoding the Adhesin Involved in *E. coli* Diffuse Adherence), *chuA* (an iron related gene), *ECs1282* (encoding a probable filamentous hemagglutinin-like protein), *rtx* (encoding a putative RTX family exoprotein), and *yjaA* (encoding a hypothetical protein) were present in all Group II strains but absent from Group I strains, including REPEC strain E22. The gene *fepC* (encoding a ferric enterobactin transport ATP-binding protein) was present in three Group II strains (ECL2019, ECL2078 and ECL2027) but absent from all other strains.

In addition, Group I and Group II strains possessed different variants of the *fliC* (encoding the flagellin major subunit) and *lpfA* genes. O45 PEPEC strains from Group I, including REPEC strain E22, possessed the *fliC* variant *flmA54*, whereas Group II strains possessed the *fliC* gene. Group I strains also possessed the *lpfA*_{O113} and *lpfA*_{R141} variants, whereas REPEC strain E22 only possessed the *lpfA*_{R141} variant, and Group II strains possessed the *lpfA1* variant.

Previous phylogenetic analyses have shown that most *E. coli* strains belonged to the four main phylogenetic groups A, B1, B2, and D. Whereas extraintestinal *E. coli* strains belong mainly to groups B2 and D, most commensal and

diarrheogenic strains belong to group A and group B1 [33]. Determination of the phylogenetic groups of the O45 PEPEC strains and REPEC strain E22 was based on the presence or absence of the two genes *chuA* and *yjaA*, and the DNA fragment *tspE4.C2*, as described by Clermont *et al.* [32]. All Group I strains including REPEC strain E22 were classified in phylogenetic group B1 and all Group II strains were classified in phylogenetic group B2.

CGH-Genomotyping of PEPEC strains

The CGH-based genomotyping analysis of the nine O45 PEPEC strains and the REPEC strain E22 led to their classification into two distinct groups in the same distribution as observed by the *E. coli* virulence microarray. A dendrogram based on the analysis of the CGH data for O45 PEPEC strains and REPEC strain E22, as well as for the two O157:H7 strains Sakai and EDL933 is presented in Figure 1. The distribution of GIs in O45 PEPEC strains and REPEC strain E22 was investigated by analysis of the CGH data. Since the microarray used for CGH was not an EPEC-specific microarray and was composed of oligonucleotide probes specific for genome sequences of O157:H7 EHEC and K12 strains, it was not possible to investigate all the GIs in O45 PEPEC strains. The divergences in GIs observed by CGH could thus indicate either the absence of particular genes or the presence of different variants of these genes. As shown in Table 2, 63 GIs (islands or loops) were found to be significantly different between Group I and Group II strains. Among these 63 GIs, 13 were KIs or

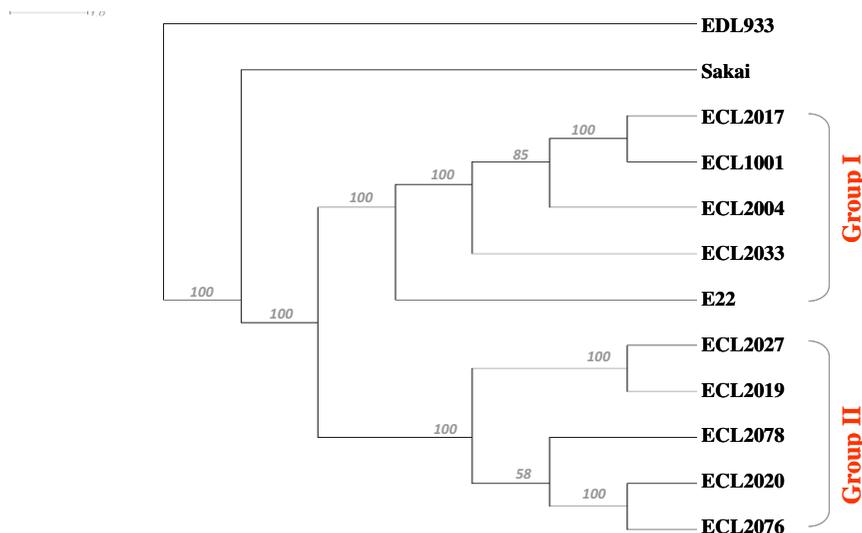


Figure 1

CGH-based genomotyping of the O45 PEPEC strains. The nine O45 PEPEC strains and the REPEC strain E22 were classified in two distinct groups by CGH-based genomotyping, in the same distribution as observed by the *E. coli* virulence microarray (Groups I and II). The O157:H7 strains Sakai and EDL933 were used as controls. The tree was constructed with tMEV v4.1 and viewed in SplitsTree 4.1 [27] by using the Euclidean distance, average linkage algorithm and 1,000 bootstrap replicates. Bootstrap confidence values are indicated at the nodes.

Table 2: Genomic islands comparison between two genotype groups of *E. coli* O45 strains.

K-Loop	K-island	O-island	S-loop	Group I ^a	Group II ^a	Notable function of this Island or Loop
		OI#1	S-loop#1	-	+	putative fimbrial chaperone and protein
		OI#2	S-loop#3	-	+	CcdA-like protein, CcdB-like protein
		OI#7	S-loop#14	+/-	-	VgrG protein
		OI#9	S-loop#17	-	+	putative transcriptional regulator, transport protein
		OI#11	S-loop#19	-	+	putative transcriptional regulator
		OI#14	S-loop#23	-	+	putative invertase, hypothetical protein
		OI#15	S-loop#24	+	-	Aida-I, adhesin-like protein
		OI#19	S-loop#30	-	+	hypothetical protein
		OI#20	S-loop#31	-	+	putative sensor histidine protein kinase
		OI#24	S-loop#37	-	+	hypothetical protein
		OI#25	S-loop#38	-	+	hypothetical protein
		OI#26	S-loop#39	-	+	hypothetical protein
		OI#28	S-loop#42	-	+	putative outer membrane transport protein
		OI#29	S-loop#43	-	+	adhesin/invasin-like protein
		OI#30	S-loop#44	+/-	-	putative Vgr protein
			S-loop#51	-	+	putative fimbrial-like protein
		OI#35	S-loop#52	-	+	putative transcription regulator
		OI#37	S-loop#57	-	+	hypothetical protein
		OI#41	S-loop#61	-	+	hypothetical protein
		OI#47	S-loop#71	-	+	hemagglutinin/hemolysin-related protein
		OI#50		+/-	-	putative antirepressor of prophage CP-933N
		OI#51	S-loop#78	+/-	-	putative single stranded DNA-binding protein
			S-loop#85	+	-	putative regulatory protein
		OI#52	S-loop#108	+	-	putative phage tail protein
		OI#52	S-loop#153	+	-	putative host specificity protein
		OI#57	S-loop#85	+	-	virulence factor Paa
		OI#65	S-loop#99	+	-	VgrE protein
		OI#70	S-loop#87	-	+	putative transcription regulatory protein
		OI#71	S-loop#108	+/-	-	BfpT-regulated chaperone-like protein
		OI#76	S-loop#119	+/-	-	putative transcriptional regulator
		OI#91		+	-	unknown function
		OI#106		-	+	putative polyferredoxin
		OI#115		-	+	TTSS or ETT2
		OI#120		+	-	hypothetical lipoprotein
		OI#122		+/-	-	putative adherence factor, putative enterotoxin
		OI#123		-	+/-	putative ABC-type iron-siderophore transport system
		OI#130		-	+	leader peptidase HopD
		OI#132		-	+	hypothetical protein
		OI#134	S-loop#220	-	+	putative ATP-dependent DNA helicase
		OI#136	S-loop#223	-	+	putative membrane protein
		OI#137	S-loop#224	-	+	hypothetical protein
		OI#140	S-loop#231	-	+	heme utilization/transport protein,
		OI#141	S-loop#232	-	+/-	putative fimbrial protein precursor
		OI#154	S-loop#253	-	+	putative type I fimbrial protein
		OI#155	S-loop#256	-	+	hypothetical protein
		OI#156	S-loop#257	+	-	hypothetical membrane protein
		OI#158	S-loop#264	-	+	hypothetical protein
		OI#159	S-loop#265	-	+	putative glycoprotein
		OI#161	S-loop#267	-	+	hypothetical membrane protein
		OI#162		+	-	RhsC protein in rhs element
K-loop#28	KI #18			+	-	polysaccharide metabolism; YaiP
K-loop#67	KI #40			+	-	hypothetical protein
K-loop#76	KI #44			+	-	hypothetical protein
K-loop#88	KI #52			+	-	homolog of virulence factor; <i>b1121</i>
K-loop#97	KI #60			+ ^b	-	orf, hypothetical protein; YchG
K-loop#140	KI #84			-	+	orf, hypothetical protein; <i>b2071</i>
K-loop#164	KI #93			-	+	acetyl-CoA: acetoacetyl-CoA transferase alpha subunit
K-loop#178	KI #103			+	-	xanthosine permease; XapA, XapB
K-loop#203	KI #120			+	-	putative DEOR-type transcriptional regulator; Ygbl

Table 2: Genomic islands comparison between two genotype groups of *E. coli* O45 strains. (Continued)

K-loop#218	KI #132	+	-	hypothetical protein
K-loop#220	KI #133	+	-	putative fimbrial-like protein; YgiL
K-loop#255	KI #154	+	-	putative lipase; YiaL
K-loop#267	KI #161	+	-	regulator protein for dgo operon; YidW

^a +/- indicates that more than 50% of the ORFs in this island or loop are present.

^b REPEC strain E22 is "-".

K-loops and 50 were OIs or S-loops. Twenty GIs were present in Group I strains but absent in Group II strains, including OI#57 (S-loop#85), which contains the *paa* gene, and the two virulence related GIs, OI#71 (S-loop#108) [34] and OI#122 [35,36]. On the other hand, 33 GIs were absent in Group I strains but present in Group II strains, including OI#1 (S-loop#1), OI#47 (S-loop#71) and OI#154 (S-loop#253), which contain fimbriae related genes. KI#60 (K-loop#97), which was absent in all Group II strains, was present in the four O45 PEPEC strains from Group I but not in REPEC strain E22. Interestingly, for seven GIs, more than half of the ORFs in each island were present in all Group I strains whereas these GIs were absent in all Group II strains. Conversely, for two GIs, more than half of the ORFs in each island were present in Group II strains whereas these GIs were absent in Group I strains.

In addition to the 63 GIs found to be significantly different between Group I and Group II strains, 26 other GIs were conserved in both Groups (Table 3). Eight of these GIs were KIs or K-loops and 18 were OIs or S-loops.

Finally, analysis of the *E. coli* O157:H7 microarray data indicated that the Shiga toxin encoding genes *stx1* and *stx2* could not be detected in any of the O45 PEPEC strains or in the REPEC strain E22, and that similarly, all strains were lacking both the *tccp* (ECs2715/Z3072) and *tccp2* (ECs1126/Z1385) genes, which encode *E. coli* O157:H7 type III effector proteins that couple the intimin receptor Tir to the actin-cytoskeleton, and trigger actin polymerization [37-40].

Analysis of LEE by the *E. coli* O157:H7 microarray

Thirty of the 41 genes on LEE (OI#148/S-loop#244) were found to be conserved among the O45 PEPEC strains,

Table 3: Conserved genomic islands identified by CGH in O45 PEPEC strains.

K-loop	K-island	O-island	S-loop	Number of ORFs	Notable function of this Island or Loop (if known)
		OI#17	S-loop #27	5	ribose transport related protein
			S-loop #45	1	
		OI #32		1	
		OI #61	S-loop #104	2	fimbrial protein
		OI #69	S-loop #89	2	
		OI #71	S-loop #109	1	chaperone protein
		OI #94		1	
		OI #103	S-loop #173	2	
		OI #108	S-loop #56	1	chaperone protein
		OI #110	S-loop #191	4	decarboxylase protein
			S-loop #192	8	
		OI#118		1	
		OI#119	S-loop #202	5	ABC transport protein
		OI#120	S-loop #203	1	
		OI#126	S-loop #211	2	sugar PTS protein
		OI#130	S-loop #215	1	bacterioferritin protein
		OI#145	S-loop #239	7	LPS biosynthesis protein
		OI#169	S-loop #278	2	
K-loop#90	KI #56			12	
K-loop#108	KI #66			20	
K-loop#115	KI #71			1	
K-loop#119	KI #73			1	
K-loop#126	KI #76			1	
K-loop#134	KI #80			1	
K-loop#232	KI #141			8	
K-loop#271	KI#163			8	

REPEC strain E22, and the two O157:H7 strains EDL933 and Sakai (Table 4). These included the effector-encoding genes *espA*, *espF* and *espG*, the regulator *ler*, and most of the genes of the type III secretion pathway such as *sepL*, *escD*, *cesT*, *escN*, *escV*, *sepD*, *escC*, *cesD*, *escU*, *escT*, *escS*, and *escR*. For the 11 remaining genes on LEE, no hybridization was observed in the O45 PEPEC strains and REPEC strain E22, possibly reflecting genetic divergences between these strains and the O157:H7 representative strains EDL933 and Sakai. These genes were the effector-encoding genes *espB*, *espD*, and *espH*, intimin and the translocated intimin receptor-encoding genes *eae* and *tir*, the genes of the type III secretion pathway *sepQ*, *sepZ*, and *escJ*, and the genes *map*, *mpc* (for multiple point controller) and Z5117.

Localization of LEE and OI#122

The LEE of AECC strains is often inserted in the vicinity of the tRNA loci *selC* or *pheU*. Since it has been previously

reported that the site of insertion of LEE in PEPEC strains could be either in *selC* or in *pheU* [23], the O45 PEPEC strains in our study and REPEC strain E22 were examined by PCR using primers specific for these two genes and for LEE extremities (Additional file 1: Table S1 [21,23,28-31]). The LEE was found to be inserted into the tRNA *pheU* locus in all examined strains. Remarkably, an amplicon of 500 bp longer than the expected size was also obtained with primers specific for LEE extremities and *selC* for strains ECL2033 and ECL2020 (Additional file 3: Table S3 [22]).

Similarly, the localization and integrity of OI#122 was determined by PCR using primers described previously (Additional file 1: Table S1 [21,23,28-31]). All Group I strains possessed this GI and were positive for the four genes tested; *efa1*, *ent*, *nleB*, and *nleE*, with the latter two encoding non-LEE virulence factors. OI#122 was found to

Table 4: Divergence in the LEE genes among O45 PEPEC strains and REPEC strain E22.

Divergent genes ^{a, b}	Group I strains ^c					Group II strains ^c					Function
	ECL1001	ECL2017	ECL2004	ECL2033	E22	ECL2019	ECL2078	ECL2027	ECL2020	ECL2076	
Z5105/ECs4554	-	+	-	-	-	-	-	-	-	-	EspB protein
Z5106/ECs4555	+	+	+	+	-	+	-	+	-	+	EspD protein
Z5110/ECs4559	-	-	-	-	-	-	-	-	-	-	Gamma intimin
Z5112/ECs4561	+	+	+	-	-	-	-	-	-	+	translocated intimin receptor Tir
Z5113/ECs4562	-	+	+	+	-	-	-	+	-	+	Map protein
Z5115/ECs4564	-	-	-	-	-	-	-	-	-	-	EspH protein
Z5116/ECs4565	-	+	+	-	+	+	-	+	-	+	type III secretion system SepQ protein
Z5117/ECs4566	-	+	+	+	-	+	-	+	-	+	hypothetical protein
Z5121/ECs4570	-	-	-	-	-	-	-	-	-	-	Mpc protein
Z5122/ECs4571	-	-	-	-	-	-	-	-	-	-	type III secretion system SepZ protein
Z5124/ECs4573	+	+	+	+	-	+	-	+	-	+	type III secretion system EscJ protein

^a Z, nomenclature for genes from O157:H7 strain EDL933; ECs, nomenclature for genes from O157:H7 strain Sakai.

^b the following genes were present in all O45 PEPEC strains and REPEC strain E22: Z5100/ECs4550 (EspF protein), Z5102/ECs4551 (hypothetical protein), Z5103/ECs4552 (EscF protein), Z5104/ECs4553 (hypothetical protein), Z5107/ECs4556 (EspA protein), Z5108/ECs4557 (type III secretion system SepL protein), Z5109/ECs4558 (type III secretion system EscD protein), Z5111/ECs4560 (CesT protein), Z5114/ECs4563 (hypothetical protein), Z5118/ECs4567 (hypothetical protein), Z5119/ECs4568 (type III secretion system protein EscN), Z5120/ECs4569 (type III secretion system EscV protein), Z5123/ECs4572 (hypothetical protein), Z5125/ECs4574 (type III secretion system protein SepD), Z5126/ECs4575 (type III secretion system EscC protein), Z5127/ECs4576 (type III secretion system CesD protein), Z5128/ECs4577 (hypothetical protein), Z5129/ECs4578 (hypothetical protein), Z5131/ECs4579 (hypothetical protein), Z5132/ECs4580 (type III secretion system EscU protein), Z5133/ECs4581 (type III secretion system EscT protein), Z5134/ECs4582 (type III secretion system EscS protein), Z5135/ECs4583 (type III secretion system EscR protein), Z5136/ECs4584 (hypothetical protein), Z5137/ECs4585 (hypothetical protein), Z5138/ECs4586 (hypothetical protein), Z5139/ECs4587 (hypothetical protein), Z5140/ECs4588 (Ler protein), Z5142/ECs4590 (EspG protein), and Z5143/ECs4591 (hypothetical protein).

^c +, gene present; -, gene absent.

be inserted into the tRNA *pheU* locus in strain ECL2033 and into the tRNA *pheV* locus in strains ECL1001 and ECL2004. The site of insertion of this GI was not determined for strain ECL2017 or for REPEC strain E22. All Group II strains lacked OI#122 (Additional file 3: Table S3 [22]).

nle genes in O45 PEPEC strains

The *E. coli* O157:H7 microarray used in our CGH studies contains oligonucleotide probes specific for genes encoding non-LEE factors which have previously been associated with the pathogenicity of AEEC strains [34,41]. The *nleA* and *nleC* genes were present in all O45 PEPEC strains and REPEC strain E22 as determined by the O157:H7

microarray (Table 5). Nevertheless, PCR analysis using two different primer sets for each gene (Additional file 1: Table S1 [21,23,28-31]) revealed amplicons of various sizes in the different strains, showing that the *nleA* and *nleC* genes in Group I were different from those in Group II (data not shown).

Sixteen other *nle* genes showed non-random distributions between Group I and Group II strains (Table 5). The five genes *espY3*, *espX2*, *espR1*, *espL3'* (Z5199/ECs4642), and *espL3'* (Z5200/ECs4643) were absent in all Group I strains but present in Group II strains. On the other hand, the five genes *espX7*, *espK*, *espL2*, *nleB1*, and *nleE* were present in all Group I strains but absent in Group II strains. The four

Table 5: Distribution of *nle* genes among O45 PEPEC strains and REPEC strain E22.

Gene	ID ^a	Family	Group I strains					Group II strains				
			ECL1001	ECL2017	ECL2004	ECL2033	E22	ECL2019	ECL2078	ECL2027	ECL2020	ECL2076
<i>espY3</i>	Z0521 (ECs0472)	SopD-N; PRR	-	-	-	-	-	+	+	+	+	+
<i>espX2</i>	Z1019 (ECs0876)	PPR	-	-	-	-	-	+	+	+	+	+
<i>espR1</i>	Z2242 (ECs2073)	LRR	-	-	-	-	-	+	+	+	+	+
<i>espL3'</i>	Z5199 (ECs4642)	AR	-	-	-	-	-	+	+	+	+	+
<i>espL3'</i>	Z5200 (ECs4643)	AR	-	-	-	-	-	+	+	+	+	+
<i>espX7</i>	Z1822 (ECs1560)	PPR; LRR	+	+	+	+	+	-	-	-	-	-
<i>espK</i>	Z1829 (ECs1568)	LRR	+	+	+	+	+	-	-	-	-	-
<i>espL2</i>	Z4326 (ECs3855)	AR	+	+	+	+	+	-	-	-	-	-
<i>nleB1</i>	Z4328 (ECs3857)	NleB	+	+	+	+	+	-	-	-	-	-
<i>nleE</i>	Z4329 (ECs3858)	NleE	+	+	+	+	+	-	-	-	-	-
<i>nleB2-1</i>	Z0985 (ECs0846)	NleB	-	-	-	-	+	+	+	+	+	+
<i>nleG2-1'</i>	Z6025 (ECs1810)	NleG	+	+	+	+	-	-	-	-	-	-
<i>espO1-2</i>	ECs1821	OspE	+	+	+	+	-	-	-	-	-	-
<i>nleG</i>	Z6010 (ECs1824)	NleG	+	+	+	+	-	-	-	-	-	-
<i>nleG9'</i>	Z2560 (ECs1828)	NleG	+	+	+	+	-	-	-	-	-	-
<i>nleD</i>	Z0990 (ECs0850)	NleD	-	-	+	+	-	-	-	-	-	-
<i>nleF</i>	Z6020 (ECs1815)	NleF	+	+	+	+	-	+	+	+	+	+
<i>nleH</i>	Z0989 (ECs0848)	NleH	+	+	+	+	+	+	+	+	+	+
<i>nleA^b</i>	Z6024 (ECs1812)	NleA	+	+	+	+	+	+	+	+	+	+
<i>nleC^b</i>	Z0986 (ECs0847)	NleC	+	+	+	+	+	+	+	+	+	+

^a Z, nomenclature for O157:H7 strain EDL933; ECs, nomenclature for O157:H7 strain Sakai.

^b PCRs using two different primer sets for each gene revealed that both *nleA* and *nleC* in Group I were different from those in Group II.

genes *nleG2-1'*, *espO1-2*, *nleG*, and *nleG9'* were present in all Group I strains with the exception of REPEC strain E22, and absent in Group II strains. The gene *nleB2-1* was present in all Group II strains and also in REPEC strain E22 but absent in the other Group I strains. The gene *nleD* was present in only two Group I strains, ECL2004 and ECL2033, and absent in all the other strains, including REPEC strain E22.

Two additional *nle* genes, *nleF* and *nleH*, were present in all O45 PEPEC strains. *nleH*, but not *nleF*, was also present in REPEC strain E22 (Table 5).

Distribution of ETT2 genes

OI#115, initially described in *E. coli* of serotype O157:H7 and present in other EPEC and EHEC strains from animals and humans, contains the secondary type III secretion system gene cluster ETT2 [31,42,43]. CGH data analysis revealed that all Group II strains had the entire ETT2 locus comprising 36 genes, with the exception of strain ECL2019, which lacked most of this GI (Additional file 4: Table S4 [16]).

In contrast, Group I strains possessed only a partially intact locus and the occurrence of the ETT2 genes was highly variable. Among the 36 ETT2 genes, strain ECL2033 possessed only 21, strain ECL2004 possessed 20, strain ECL1001 possessed 17, and strain ECL2017 possessed 15. Finally, REPEC strain E22 possessed 21 genes of this GI. These results were confirmed by PCR as described previously [31], with primers specific for different regions of the ETT2 gene cluster (Additional file 1: Table S1 [21,23,28-31]).

Genes required for intestinal colonization in the bovine

Previous studies have identified several genes required for EHEC intestinal colonization of the bovine [44,45]. Microarray analysis in our study revealed that 13 genes associated with colonization of either *E. coli* O157:H7 or *E. coli* O26:H- in the bovine were associated with either Group I or Group II strains (Additional file 5: Table S5). Seven genes were present in Group I but not in Group II strains, with the exception of REPEC strain E22 which did not possess the gene Z6010 (ECs1824). In contrast, six other genes were present in Group II but not in Group I strains, with the exception of REPEC strain E22 which possessed the gene Z1526 (ECs1270). These results were confirmed by PCR using primers designed for each gene (data not shown).

Discussion

In this study, we investigated the genetic relationships among PEPEC strains of serogroup O45 and catalogued genomic alterations unique to these strains by using both a virulence gene-specific microarray and a whole genome microarray. The O45 PEPEC strains in this study have been

previously characterized for their capacity to induce A/E lesions in both explants and challenged pigs, and were grouped according to the severity of the A/E manifestation they produced [13]. Based on their virulence gene content as determined by the *E. coli* virulence microarray, O45 PEPEC strains and REPEC strain E22 displayed significant differences from typical EPEC and could be regarded as atypical EPEC, that are defined as LEE-positive *E. coli* lacking *stx*₁ and *stx*₂ genes, as well as the EAF virulence plasmid which encodes the EPEC adhesin BFP [46,47]. In addition, all O45 PEPEC strains and REPEC strain E22 unexpectedly hybridized with the *stxB*₁ probe of the *E. coli* virulence microarray, as was also observed for some atypical EPEC strains isolated from children with diarrhea in a recent study in Norway [35,48]. Due to the absence of hybridization with the corresponding *stxA*₁ probe and the negative PCR results obtained with *stxB*₁ sequence specific primers [35,48], we therefore concluded that the gene sequences detected by the *stxB*₁ hybridization probe did not represent a complete *stxB* gene but rather a possible truncated form of this gene.

As observed for other atypical EPEC strains, O45 PEPEC strains and REPEC strain E22 also displayed a relatively high heterogeneity in their virulence gene profiles [35,49]. Based on their virulence gene content, they could be divided into two distinct groups, Groups I and II. It has been argued that atypical EPEC strains could have arisen from *E. coli* strains of different pathotypes which acquired the LEE by horizontal gene transfer or from certain typical EPEC strains that have lost the EAF plasmid [49]. Trabulsi *et al.* have also observed that some atypical EPEC strains are genetically closer to EHEC strains of serotype O157:H7 than to typical EPEC [50]. Several virulence genes showed a non-random distribution between Group I and Group II strains. Group I strains thus possessed several virulence-related genes which were absent in Group II strains. Group I-specific genes included *paa* (which contributes to A/E lesion formation in PEPEC strains [10]) and OI#122 genes *efa1* (which plays an important role in intestinal colonization by EHEC strains in cattle [51]) and *set* (which encodes a putative enterotoxin highly similar to the enterotoxin ShET2 of *Shigella flexneri*). Genes associated with other pathotypes were also found. The gene *tsh*, encoding a hemagglutinin which may be a virulence factor of avian extraintestinal *E. coli* [52], the pathogenicity island marker *malX*, related to virulence in extraintestinal *E. coli* [53] and the yersiniabactin-related genes *fyuA*, *irp1* and *irp2*, implicated in the ferric uptake system, were also Group I-specific. In contrast, Group II strains possessed only a few additional virulence-related genes when compared with Group I strains. These included *aidal*, which encodes a protein involved in the adherence of EPEC [54], and iron uptake-related genes *chuA* and *fepC*. Finally, Group I and Group II strains also possessed different variants of the long polar fimbriae encoding gene *lpfA*.

A recent study has shown that the *lpfA*_{O113} variant, found in Group I strains in our study, was found significantly more frequently in atypical EPEC strains associated with cases of diarrhea than in strains isolated from healthy individuals [35].

It is interesting to note that analysis by CGH using a whole genome *E. coli* microarray, representing two lineage I, human outbreak-related *E. coli* O157:H7 strains and one non-pathogenic *E. coli* K12 strain, resulted in the distribution of the O45 PEPEC strains into the same two groups (Groups I and II), observed for the *E. coli* virulence microarray. This genetic-based grouping, principally reflecting the virulence gene content of the strains, was also compatible with the grouping based on their A/E activity [13]. The O45 PEPEC strains of Group I all induced severe A/E lesions whereas those of Group II induced less severe or no A/E lesions in both pig ileal explants and challenged pigs. REPEC strain E22 was placed into Group I but was genetically distant from the four O45 PEPEC strains belonging to this group. These strains showed a relatively high level of heterogeneity in their virulence gene profiles.

In addition to the variations in particular virulence genes, significant variations in GIs were also observed between Group I and II strains. We observed that several virulence-related OIs were present only in Group I strains. These included OI#57 (S-loop#85), which contains the *paa* gene; OI#71 (S-loop#108), which contains the non-LEE encoded factor gene *nleA*, previously shown to be associated with the pathogenicity of AEEC strains [34]; and OI#122 which contains the two non-LEE encoded factor genes *nleB* and *nleE* and virulence genes *efa1* and *set*. Interestingly, these OIs have also been shown to be more prevalent in STEC strains associated with outbreaks and severe disease [35,36]. On the other hand, certain OIs were only present in Group II strains. These include OI#1 (S-loop#1), containing genes encoding putative fimbrial chaperone proteins; OI#47 (S-loop#71), containing a fimbrial operon and genes encoding several additional putative virulence factors in *E. coli* of serogroup O157 [55]; and OI#154 (S-loop#253), containing genes encoding putative type 1 fimbrial proteins. Finally, OI#115 was highly divergent between Group I and Group II strains. This OI contains the secondary type III secretion system gene cluster ETT2, which resembles the SPI-1-encoded type III secretion system from *Salmonella enterica* and has been previously characterized in O157:H7 *E. coli* strains [16,17]. It has been recently shown that ETT2 influences the secretion of proteins encoded by the LEE and also modulates adhesion to human intestinal cells [56]. Most Group II strains (4/5) possessed the entire ETT2. However, Group II strain ECL2019 lacked most of the entire cluster (only four genes were found to be present) and was one of the two strains which did not induce A/E lesions in

explants or in challenged pigs. Group I strains, including REPEC strain E22, have only a partial ETT2 gene cluster, possessing from 15 to 21 genes of the 36 in the intact cluster. The substantial variations observed for this cluster are consistent with the findings of previous studies, that while the ETT2 gene cluster was present in most of the *E. coli* strains tested, it contained numerous inactivating mutations [31,42,43].

In contrast to the heterogeneity of their virulence gene content and GI distribution, O45 PEPEC strains and REPEC strain E22 showed a high level of homogeneity in their LEE sequences and site of insertion. In all strains, the LEE was inserted into the tRNA *pheU* gene and no significant divergence between Group I and Group II strains was observed for the LEE genes. In addition, all O45 PEPEC strains and REPEC strain E22 shared the same profile for the intimin encoding gene *eae*, its translocated receptor *tir* and the effector encoding genes *espA* and *espB*, as shown by the *E. coli* virulence microarray. All strains possessed the beta variant of the intimin encoding gene, being the most widespread among the intestinal EPEC strains of different animal species [57,58]. However, Group I and Group II strains from our study all belonged to the phylogenetic groups B1 and B2 whereas Ishii *et al.* have shown that most EPEC strains possessing the intimin subtype beta belong to phylogenetic groups A and B1 [59].

Finally, we have also observed numerous disparities in the distribution of non-LEE encoded genes. Several *nle* genes were only present in Group I strains whereas others were only found in Group II strains. In addition, Group I and Group II strains possessed two different variants of the genes *nleA* and *nleC*. This variation in the distribution of *nle* genes may influence the pathogenicity of the strains and the type of A/E lesions they produce, since many studies have shown the importance of non-LEE encoded factors in the A/E phenotype [34,41,60,61].

Conclusion

We have genetically characterized a collection of O45 PEPEC strains using *E. coli* O157-*E. coli* K12 whole genome and virulence gene-specific *E. coli* microarrays. We have shown that the strains, although showing some heterogeneity, could be classified into two groups, based on their virulence gene and GI content. These differences in their virulence gene content may influence the pathogenicity of O45 PEPEC strains, and explain why Group I O45 PEPEC strains induced more severe A/E lesions in explants and challenged pigs than Group II strains [13,14].

Authors' contributions

GB and YZ both contributed equally to the manuscript: they were involved in the conception and design of the

study, in the analysis and interpretation of the microarray data, and in drafting and revising the manuscript. PG, JW, and CL were involved in the acquisition of the microarray data and in revising the manuscript. JMF was involved in revising the manuscript critically for important intellectual content. VPJG and JH were involved in the conception and design of the study, and revising the manuscript critically for important intellectual content. All authors read and approved the final manuscript.

Additional material

Additional file 1

Table S1. Primers and *E. coli* control strains used for PCR experiments.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-402-S1.pdf>]

Additional file 2

Table S2. Presence of virulence genes in O45 PEPEC strains and REPEC strain E22 as determined by *E. coli* virulence microarray.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-402-S2.pdf>]

Additional file 3

Table S3. Localization of the *LEE* and *OI#122* in O45 PEPEC strains and REPEC strain E22.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-402-S3.pdf>]

Additional file 4

Table S4. Distribution of *ETT2* genes in O45 PEPEC strains and REPEC strain E22.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-402-S4.pdf>]

Additional file 5

Table S5. Divergence of genes related to intestinal colonization in bovine and calves as determined by CGH.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-402-S5.pdf>]

Acknowledgements

Clarisse Desautels, from the Reference Laboratory for *E. coli*, Faculté de médecine vétérinaire, Université de Montréal, is greatly acknowledged for her information and her help regarding the collection of O45 PEPEC strains. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) (STPGP 364950), and by the Fond Québécois de la Recherche sur la Nature et les Technologies (FQRNT) (PR-121927). G.B. was supported by a scholarship "Michel Saucier" from Fondation Canadienne Louis Pasteur (FCLP).

References

- Ott M, Bender L, Blum G, Schmittroth M, Achtman M, Tschape H, Hacker J: **Virulence patterns and long-range genetic mapping of extraintestinal *Escherichia coli* K1, K5, and K100 isolates: use of pulsed-field gel electrophoresis.** *Infect Immun* 1991, **59(8)**:2664-2672.
- Plainvert C, Bidet P, Peigne C, Barbe V, Medigue C, Denamur E, Bingen E, Bonacorsi S: **A new O-antigen gene cluster has a key role in the virulence of the *Escherichia coli* meningitis clone O45:K1:H7.** *J Bacteriol* 2007, **189(23)**:8528-8536.
- Wullenweber M, Beutin L, Zimmermann S, Jonas C: **Influence of some bacterial and host factors on colonization and invasiveness of *Escherichia coli* K1 in neonatal rats.** *Infect Immun* 1993, **61(5)**:2138-2144.
- Blanco J, Blanco M, Garabal JI, Gonzalez EA: **Enterotoxins, colonization factors and serotypes of enterotoxigenic *Escherichia coli* from humans and animals.** *Microbiologia* 1991, **7(2)**:57-73.
- Nagy B, Fekete PZ: **Enterotoxigenic *Escherichia coli* (ETEC) in farm animals.** *Vet Res* 1999, **30(2-3)**:259-284.
- Toth I, Karcagi V, Nagy B, Gado I, Milch H: **Examination of verocytotoxin producing capacity and determination of the presence of Shiga-like toxin genes in human *Escherichia coli* isolates.** *Acta Microbiol Immunol Hung* 1994, **41(3)**:259-264.
- Frydendahl K: **Prevalence of serogroups and virulence genes in *Escherichia coli* associated with postweaning diarrhoea and edema disease in pigs and a comparison of diagnostic approaches.** *Vet Microbiol* 2002, **85(2)**:169-182.
- Harel J, Lapointe H, Fallara A, Lortie LA, Bigras-Poulin M, Lariviere S, Fairbrother JM: **Detection of genes for fimbrial antigens and enterotoxins associated with *Escherichia coli* serogroups isolated from pigs with diarrhea.** *J Clin Microbiol* 1991, **29(4)**:745-752.
- Helie P, Morin M, Jacques M, Fairbrother JM: **Experimental infection of newborn pigs with an attaching and effacing *Escherichia coli* O45:K"E65" strain.** *Infect Immun* 1991, **59(3)**:814-821.
- Batissou I, Guimond MP, Girard F, An H, Zhu C, Oswald E, Fairbrother JM, Jacques M, Harel J: **Characterization of the novel factor *paa* involved in the early steps of the adhesion mechanism of attaching and effacing *Escherichia coli*.** *Infect Immun* 2003, **71(8)**:4516-4525.
- Girard F, Batissou I, Frankel GM, Harel J, Fairbrother JM: **Interaction of enteropathogenic and Shiga toxin-producing *Escherichia coli* and porcine intestinal mucosa: role of intimin and Tir in adherence.** *Infect Immun* 2005, **73(9)**:6005-6016.
- Girard F, Oswald IP, Taranu I, Helie P, Appleyard GD, Harel J, Fairbrother JM: **Host immune status influences the development of attaching and effacing lesions in weaned pigs.** *Infect Immun* 2005, **73(9)**:5514-5523.
- Zhu C, Harel J, Jacques M, Desautels C, Donnenberg MS, Beaudry M, Fairbrother JM: **Virulence properties and attaching-effacing activity of *Escherichia coli* O45 from swine postweaning diarrhea.** *Infect Immun* 1994, **62(10)**:4153-4159.
- Zhu C, Harel J, Jacques M, Fairbrother JM: **Interaction with pig ileal explants of *Escherichia coli* O45 isolates from swine with postweaning diarrhea.** *Can J Vet Res* 1995, **59(2)**:118-123.
- Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y: **The complete genome sequence of *Escherichia coli* K-12.** *Science* 1997, **277(5331)**:1453-1474.
- Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han CG, Ohtsubo E, Nakayama K, Murata T, Tanaka M, Tobe T, Iida T, Takami H, Honda T, Sasakawa C, Ogasawara N, Yasunaga T, Kuhara S, Shiba T, Hattori M, Shinagawa H: **Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12.** *DNA Res* 2001, **8(1)**:11-22.
- Perna NT, Plunkett G 3rd, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, Posfai G, Hackett J, Klink S, Boutin A, Shao Y, Miller L, Grotbeck EJ, Davis NW, Lim A, Dimalanta ET, Potamousis KD, Apodaca J, Anantharaman TS, Lin J, Yen G, Schwartz DC, Welch RA, Blattner FR: **Genome sequence**

- of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 2001, **409**(6819):529-533.
18. Kaper JB, Mellies JL, Nataro JP: **Pathogenicity islands and other mobile genetic elements of diarrheagenic *E. coli***. Washington D. C.: ASM Press; 1999.
 19. Zhang Y, Laing C, Steele M, Ziebell K, Johnson R, Benson AK, Taboada E, Gannon VP: **Genome evolution in major *Escherichia coli* O157:H7 lineages**. *BMC Genomics* 2007, **8**:121.
 20. Elliott SJ, Wainwright LA, McDaniel TK, Jarvis KG, Deng YK, Lai LC, McNamara BP, Donnenberg MS, Kaper JB: **The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69**. *Mol Microbiol* 1998, **28**(1):1-4.
 21. Sperandio V, Kaper JB, Bortolini MR, Neves BC, Keller R, Trabulsi LR: **Characterization of the locus of enterocyte effacement (LEE) in different enteropathogenic *Escherichia coli* (EPEC) and Shiga-toxin producing *Escherichia coli* (STEC) serotypes**. *FEMS Microbiol Lett* 1998, **164**(1):133-139.
 22. Bruant G, Maynard C, Bekal S, Gaucher I, Masson L, Brousseau R, Harel J: **Development and validation of an oligonucleotide microarray for detection of multiple virulence and antimicrobial resistance genes in *Escherichia coli***. *Appl Environ Microbiol* 2006, **72**(5):3780-3784.
 23. An H, Fairbrother JM, Desautels C, Mabrouk T, Dugourd D, Dezfulian H, Harel J: **Presence of the LEE (locus of enterocyte effacement) in pig attaching and effacing *Escherichia coli* and characterization of *eae*, *espA*, *espB* and *espD* genes of PEPEC (pig EPEC) strain 1390**. *Microb Pathog* 2000, **28**(5):291-300.
 24. Beutin L, Kaulfuss S, Herold S, Oswald E, Schmidt H: **Genetic analysis of enteropathogenic and enterohemorrhagic *Escherichia coli* serogroup O103 strains by molecular typing of virulence and housekeeping genes and pulsed-field gel electrophoresis**. *J Clin Microbiol* 2005, **43**(4):1552-1563.
 25. Camguilhem R, Milon A: **Biotypes and O serogroups of *Escherichia coli* involved in intestinal infections of weaned rabbits: clues to diagnosis of pathogenic strains**. *J Clin Microbiol* 1989, **27**(4):743-747.
 26. Kim CC, Joyce EA, Chan K, Falkow S: **Improved analytical methods for microarray-based genome-composition analysis**. *Genome Biol* 2002, **3**(11):RESEARCH0065.
 27. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J: **TM4: a free, open-source system for microarray data management and analysis**. *Biotechniques* 2003, **34**(2):374-378.
 28. McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB: **A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens**. *Proc Natl Acad Sci USA* 1995, **92**(5):1664-1668.
 29. Bielaszewska M, Sonntag AK, Schmidt MA, Karch H: **Presence of virulence and fitness gene modules of enterohemorrhagic *Escherichia coli* in atypical enteropathogenic *Escherichia coli* O26**. *Microbes Infect* 2007, **9**(7):891-897.
 30. Bertin Y, Boukhors K, Livrelli V, Martin C: **Localization of the insertion site and pathotype determination of the locus of enterocyte effacement of shiga toxin-producing *Escherichia coli* strains**. *Appl Environ Microbiol* 2004, **70**(1):61-68.
 31. Hartleib S, Prager R, Hedenstrom I, Lofdahl S, Tschape H: **Prevalence of the new, SPII-like, pathogenicity island ETT2 among *Escherichia coli***. *Int J Med Microbiol* 2003, **292**(7-8):487-493.
 32. Clermont O, Bonacorsi S, Bingen E: **Rapid and simple determination of the *Escherichia coli* phylogenetic group**. *Appl Environ Microbiol* 2000, **66**(10):4555-4558.
 33. Girardeau JP, Dalmaso A, Bertin Y, Ducrot C, Bord S, Livrelli V, Vernozy-Rozand C, Martin C: **Association of virulence genotype with phylogenetic background in comparison to different seropathotypes of Shiga toxin-producing *Escherichia coli* isolates**. *J Clin Microbiol* 2005, **43**(12):6098-6107.
 34. Gruenheid S, Sekirov I, Thomas NA, Deng W, O'Donnell P, Goode D, Li Y, Frey EA, Brown NF, Metalnikov P, Pawson T, Ashman K, Finlay BB: **Identification and characterization of NleA, a non-LEE-encoded type III translocated virulence factor of enterohaemorrhagic *Escherichia coli* O157:H7**. *Mol Microbiol* 2004, **51**(5):1233-1249.
 35. Afset JE, Bruant G, Brousseau R, Harel J, Anderssen E, Bevanger L, Bergh K: **Identification of virulence genes linked with diarrhea due to atypical enteropathogenic *Escherichia coli* by DNA microarray analysis and PCR**. *J Clin Microbiol* 2006, **44**(10):3703-3711.
 36. Karmali MA, Mascarenhas M, Shen S, Ziebell K, Johnson S, Reid-Smith R, Isaac-Renton J, Clark C, Rahn K, Kaper JB: **Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease**. *J Clin Microbiol* 2003, **41**(11):4930-4940.
 37. Frankel G, Phillips AD: **Attaching effacing *Escherichia coli* and paradigms of Tir-triggered actin polymerization: getting off the pedestal**. *Cell Microbiol* 2008, **10**(3):549-556.
 38. Ogura Y, Ooka T, Whale A, Garmendia J, Beutin L, Tennant S, Krause G, Morabito S, Chinen I, Tobe T, Abe H, Tozzoli R, Caprioli A, Rivas M, Robins-Browne R, Hayashi T, Frankel G: **TccP2 of O157:H7 and non-O157 enterohemorrhagic *Escherichia coli* (EHEC): challenging the dogma of EHEC-induced actin polymerization**. *Infect Immun* 2007, **75**(2):604-612.
 39. Whale AD, Garmendia J, Gomes TA, Frankel G: **A novel category of enteropathogenic *Escherichia coli* simultaneously utilizes the Nck and TccP pathways to induce actin remodelling**. *Cell Microbiol* 2006, **8**(6):999-1008.
 40. Whale AD, Hernandez RT, Ooka T, Beutin L, Schuller S, Garmendia J, Crowther L, Vieira MA, Ogura Y, Krause G, Phillips AD, Gomes TA, Hayashi T, Frankel G: **TccP2-mediated subversion of actin dynamics by EPEC 2 - a distinct evolutionary lineage of enteropathogenic *Escherichia coli***. *Microbiology* 2007, **153**(Pt 6):1743-1755.
 41. Deng W, Puente JL, Gruenheid S, Li Y, Vallance BA, Vazquez A, Barba J, Ibarra JA, O'Donnell P, Metalnikov P, Ashman K, Lee S, Goode D, Pawson T, Finlay BB: **Dissecting virulence: systematic and functional analyses of a pathogenicity island**. *Proc Natl Acad Sci USA* 2004, **101**(10):3597-3602.
 42. Makino S, Tobe T, Asakura H, Watarai M, Ikeda T, Takeshi K, Sasaki K: **Distribution of the secondary type III secretion system locus found in enterohemorrhagic *Escherichia coli* O157:H7 isolates among Shiga toxin-producing *E. coli* strains**. *J Clin Microbiol* 2003, **41**(6):2341-2347.
 43. Ren CP, Chaudhuri RR, Fivian A, Bailey CM, Antonio M, Barnes WM, Pallen MJ: **The ETT2 gene cluster, encoding a second type III secretion system from *Escherichia coli*, is present in the majority of strains but has undergone widespread mutational attrition**. *J Bacteriol* 2004, **186**(11):3547-3560.
 44. Dziva F, van Diemen PM, Stevens MP, Smith AJ, Wallis TS: **Identification of *Escherichia coli* O157:H7 genes influencing colonization of the bovine gastrointestinal tract using signature-tagged mutagenesis**. *Microbiology* 2004, **150**(Pt 11):3631-3645.
 45. van Diemen PM, Dziva F, Stevens MP, Wallis TS: **Identification of enterohemorrhagic *Escherichia coli* O26:H- genes required for intestinal colonization in calves**. *Infect Immun* 2005, **73**(3):1735-1743.
 46. Kaper JB: **Defining EPEC**. *Rev Microbiol Sao Paulo* 1996, **27**:130-133.
 47. Malik A, Toth I, Beutin L, Schmidt H, Taminiau B, Dow MA, Morabito S, Oswald E, Mainil J, Nagy B: **Serotypes and intimin types of intestinal and faecal strains of *eae* + *Escherichia coli* from weaned pigs**. *Vet Microbiol* 2006, **114**(1-2):82-93.
 48. Afset JE, Anderssen E, Bruant G, Harel J, Wieler L, Bergh K: **Phylogenetic backgrounds and virulence profiles of atypical enteropathogenic *Escherichia coli* strains from a case-control study using multilocus sequence typing and DNA microarray analysis**. *J Clin Microbiol* 2008, **46**(7):2280-2290.
 49. Vieira MA, Andrade JR, Trabulsi LR, Rosa AC, Dias AM, Ramos SR, Frankel G, Gomes TA: **Phenotypic and genotypic characteristics of *Escherichia coli* strains of non-enteropathogenic *E. coli* (EPEC) serogroups that carry EAE and lack the EPEC adherence factor and Shiga toxin DNA probe sequences**. *J Infect Dis* 2001, **183**(5):762-772.
 50. Trabulsi LR, Keller R, Tardelli Gomes TA: **Typical and atypical enteropathogenic *Escherichia coli***. *Emerg Infect Dis* 2002, **8**(5):508-513.
 51. Nicholls L, Grant TH, Robins-Browne RM: **Identification of a novel genetic locus that is required for in vitro adhesion of a clinical isolate of enterohaemorrhagic *Escherichia coli* to epithelial cells**. *Mol Microbiol* 2000, **35**(2):275-288.

52. Dozois CM, Dho-Moulin M, Bree A, Fairbrother JM, Desautels C, Curtiss R 3rd: **Relationship between the Tsh autotransporter and pathogenicity of avian *Escherichia coli* and localization and analysis of the Tsh genetic region.** *Infect Immun* 2000, **68(7)**:4145-4154.
53. Johnson JR, Stell AL: **Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise.** *J Infect Dis* 2000, **181(1)**:261-272.
54. Benz I, Schmidt MA: **Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic *Escherichia coli*.** *Infect Immun* 1989, **57(5)**:1506-1511.
55. Shen S, Mascarenhas M, Morgan R, Rahn K, Karmali MA: **Identification of four fimbria-encoding genomic islands that are highly specific for verocytotoxin-producing *Escherichia coli* serotype O157 strains.** *J Clin Microbiol* 2005, **43(8)**:3840-3850.
56. Zhang L, Chaudhuri RR, Constantinidou C, Hobman JL, Patel MD, Jones AC, Sarti D, Roe AJ, Vlisidou I, Shaw RK, Falciani F, Stevens MP, Gally DL, Knutton S, Frankel G, Penn CV, Pallen MJ: **Regulators encoded in the *Escherichia coli* type III secretion system 2 gene cluster influence expression of genes within the locus for enterocyte effacement in enterohemorrhagic *E. coli* O157:H7.** *Infect Immun* 2004, **72(12)**:7282-7293.
57. Oswald E, Schmidt H, Morabito S, Karch H, Marches O, Caprioli A: **Typing of intimin genes in human and animal enterohemorrhagic and enteropathogenic *Escherichia coli*: characterization of a new intimin variant.** *Infect Immun* 2000, **68(1)**:64-71.
58. Krause G, Zimmermann S, Beutin L: **Investigation of domestic animals and pets as a reservoir for intimin- (*eae*) gene positive *Escherichia coli* types.** *Vet Microbiol* 2005, **106(1-2)**:87-95.
59. Ishii S, Meyer KP, Sadowsky MJ: **Relationship between phylogenetic groups, genotypic clusters, and virulence gene profiles of *Escherichia coli* strains from diverse human and animal sources.** *Appl Environ Microbiol* 2007, **73(18)**:5703-5710.
60. Kelly M, Hart E, Mundy R, Marches O, Wiles S, Badea L, Luck S, Tauschek M, Frankel G, Robins-Browne RM, Hartland EL: **Essential role of the type III secretion system effector NleB in colonization of mice by *Citrobacter rodentium*.** *Infect Immun* 2006, **74(4)**:2328-2337.
61. Roe AJ, Tysall L, Dransfield T, Wang D, Fraser-Pitt D, Mahajan A, Constantinou C, Inglis N, Downing A, Talbot R, Smith DG, Gally DL: **Analysis of the expression, regulation and export of NleA-E in *Escherichia coli* O157:H7.** *Microbiology* 2007, **153(Pt 5)**:1350-1360.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

