Comparative genomics of Crohn's disease-associated adherent-invasive Escherichia coli

Claire L O'Brien, Marie-Agnès Bringer, Kathryn E Holt, David M Gordon, Anaëlle L Dubois, Nicolas Barnich, Arlette Darfeuille-Michaud, Paul Pavli

Gut Online First, published on May 19, 2016 as 10.1136/gutjnl-2015-311059

OPEN ACCESS

ORIGINAL ARTICLE

Comparative genomics of Crohn's disease-associated adherent-invasive *Escherichia coli*

Claire L O'Brien,[1,2] Marie-Agnès Bringer,[3,4] Kathryn E Holt,[5] David M Gordon,[6] Anaëlle L Dubois,[4] Nicolas Barnich,[4] Arlette Darfeuille-Michaud,[4] Paul Pavli[1,2]

▸ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2015-311059).

[1]Medical School, Australian National University, Canberra, Australian Capital Territory, Australia
[2]Gastroenterology and Hepatology Unit, Canberra Hospital, Canberra, Australian Capital Territory, Australia
[3]INRA UMR1324, CNRS UMR6265, Université Bourgogne-Franche-Comté, Centre des Sciences du Goût et de l'Alimentation, Dijon, France
[4]UMR1071 Inserm/University of Auvergne, INRA USC2018, M2iSH, Clermont-Ferrand, France
[5]Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Melbourne, Victoria, Australia
[6]Research School of Biology, Australian National University, Canberra, Australian Capital Territory, Australia

Correspondence to
Dr Claire O'Brien, Medical School, Australian National University, Canberra Hospital, Lvl 5, bldg. 10, Yamba Drive, Garran, ACT 2605, Australia; claire.obrien@anu.edu.au

Received 8 November 2015
Accepted 22 March 2016

ABSTRACT

Objective Adherent-invasive *Escherichia coli* (AIEC) are a leading candidate bacterial trigger for Crohn's disease (CD). The AIEC pathovar is defined by in vitro cell-line assays examining specific bacteria/cell interactions. No molecular marker exists for their identification. Our aim was to identify a molecular property common to the AIEC phenotype.

Design 41 B2 phylogroup *E. coli* strains were isolated from 36 Australian subjects: 19 patients with IBD and 17 without. Adherence/invasion assays were conducted using the I-407 epithelial cell line and survival/replication assays using the THP-1 macrophage cell line. Cytokine secretion tumour necrosis factor ((TNF)-α, interleukin (IL) 6, IL-8 and IL-10) was measured using ELISA. The genomes were assembled and annotated, and cluster analysis performed using CD-HIT. The resulting matrices were analysed to identify genes unique/more frequent in AIEC strains compared with non-AIEC strains. Base composition differences and clustered regularly interspaced palindromic repeat (CRISPR) analyses were conducted.

Results Of all B2 phylogroup strains assessed, 79% could survive and replicate in macrophages. Among them, 11/41 strains (5 CD, 2 UCs, 5 non-IBD) also adhere to and invade epithelial cells, a phenotype assigning them to the AIEC pathovar. The AIEC strains were phylogenetically heterogeneous. We did not identify a gene (or nucleic acid base composition differences) common to all, or the majority of, AIEC. Cytokine secretion and CRISPRs were not associated with the AIEC phenotype.

Conclusions Comparative genomic analysis of AIEC and non-AIEC strains did not identify a molecular property exclusive to the AIEC phenotype. We recommend a broader approach to the identification of the bacteria-host interactions that are important in the pathogenesis of Crohn's disease.

Significance of this study

What is already known on this subject?

- Identifying the functional and/or genetic properties of bacterial triggers of Crohn's disease may lead to effective therapeutic strategies.
- Adherent-invasive *Escherichia coli* (AIEC) are more commonly isolated from mucosal biopsies of patients with Crohn's disease than controls.
- AIEC are defined by their phenotype: the ability, in vitro, to adhere to and invade epithelial cell lines and survive and replicate within macrophages.
- To date, no virulence factor, gene or combination of genes, has been found to explain the AIEC phenotype.

What are the new findings?

- Whole-genome sequencing could not define a specific genotype that explains the AIEC phenotype.
- Survival and replication within macrophages is a common feature of *E. coli* isolated from intestinal mucosa generally.

How might it impact on clinical practice in the foreseeable future?

- Beyond AIEC, other aspects of the host-microbiome interaction should be examined.

INTRODUCTION

Crohn's disease (CD) is a complex disease that is thought to result from interactions between luminal microbes and the host innate immune system, in genetically susceptible individuals.[1] The number of host susceptibility loci identified for the IBDs continues to grow: over 200 have now been identified, 30 of which are CD-specific.[2 3] The most significant, replicable host mutations encode genes involved in the detection of, signalling in response to and clearance of, bacteria. The gut microbiome is altered in patients with CD relative to controls:[4–8] studies show an increase in mucosa-associated *Escherichia coli* in both ileum and colon,[4 9–14] but no single causative micro-organism has been identified.

The adherent-invasive *E. coli* (AIEC) pathovar was first described by Boudeau *et al*[15] in 1999, and has emerged as a leading candidate bacterial trigger for CD. *E. coli* belonging to this pathovar are defined by their in vitro abilities to adhere to and to invade epithelial cells, and to survive and replicate within macrophages.[9 16 17]

E. coli is a phylogenetically diverse species, and strains are often assigned one of eight major phylogroups (A, B1, B2, C, D, E, F and *Escherichia* Cryptic Clade i) by PCR.[18] Approximately 47% of strains isolated from human mucosal biopsies are phylogroup B2, which, when present in a host, are likely to be the dominant strain.[19] The majority

To cite: O'Brien CL, Bringer M-A, Holt KE, *et al*. *Gut* Published Online First: [*please include* Day Month Year] doi:10.1136/gutjnl-2015-311059

(64%) of AIEC isolates belong to the B2 phylogroup,[12] and this phylogroup typically contains more virulence factors than strains belonging to other phylogroups.[20] Multilocus sequence typing (MLST) is also used to assign *E. coli* into sequence types (STs), using one of three schemes.[21–23] Any mucosa-associated strain, regardless of phylogroup, is likely to be found along the length of the lower GI tract, and is unlikely to be restricted to a single region.[19] ST95 strains are commonly isolated from humans and belong to the B2 phylogroup.[24]

AIEC are more commonly isolated from mucosal biopsies of patients with CD (36–52%) than controls (6–17%),[9 12] and one study showed that only 6.3% of extraintestinal pathogenic *E. coli* (ExPEC) strains were AIEC.[25] As observed for other pathogens, such as uropathogenic *E. coli* (UPEC), different mechanisms are involved in the colonisation of the epithelium by AIEC. For example, it is thought that AIEC colonise the ileal mucosa in patients with CD through abnormal expression of carcinoembryonic antigen-related cell adhesion molecule 6 (CEACAM6) receptors recognising type 1 pili.[26 27] Point mutations in the *FimH* adhesion of AIEC result in an increased ability to adhere to CEACAM6-expressing intestinal epithelial cells.[28] AIEC also target M cells on Peyer's patches through the expression of type 1 pili and long polar fimbriae (LPF), a mechanism allowing them to translocate across the intestinal epithelial barrier.

Several studies have attempted to identify virulence factors associated with the AIEC phenotype. Conte *et al*[17] showed that some virulence genes are more frequent in the *E. coli* isolates from patients with CD than controls, including: K1 and *kpsMT* II, both involved in capsule synthesis; *fyuA*, involved in iron acquisition; and *ibeA*, involved in invasion. The relative abundance of these strains was significantly higher in patients with CD (10%) compared with controls (1%). Comparative genome sequencing conducted by Dogan *et al*[29] on isolates from different origins (patients with CD, dogs with granulomatous colitis and mouse ileitis) failed to detect a molecular property associated with the AIEC phenotype. However, their study revealed that certain factors were associated with CD-derived AIEC, including *pduC* (a putative glycerol dehydatase) and *chuA*, (involved in haem acquisition), which is present in all B2 strains of *E. coli*. CD-associated AIEC harbouring *lpfA* (involved in cell attachment) displayed a high level of invasion of epithelial cells and translocation through M cells. Desilets *et al*[30] conducted comparative genome analyses on a panel of *E. coli* strains, containing 14 IBD-associated strains that were only assessed for their ability to replicate intracellularly in the RAW264.7 macrophage cell line, and 40 published genomes comprising various pathovars of *E. coli* including AIEC. They did not identify a gene in common to all AIEC/IBD-associated *E. coli* strains, but suggested that B2 phylogroup AIEC may represent a distinct cluster of IBD-associated *E. coli*. Recently, Deshpande *et al*[31] conducted a genome comparison of four AIEC and five other *E. coli* strains belonging to the UPEC, ExPEC and Avian pathogenic Escherichia coli (APEC) pathovars. They identified six amino acid changes associated with all nine strains, then used these amino acid changes to scan a large set of *E. coli* strains. Of the 1311 strains, 73 clustered with the 9 original strains. Because seven of these were ST95, the majority of strains that clustered with the AIEC strains were also ST95. It is likely that the associations they describe are phylogenetic in nature and do not reflect the pathogenic potential of the strains.

Currently, the only way to identify AIEC strains is by conducting bacteria/cell interaction assays. Although several molecular markers have been associated with AIEC and/or play a role in AIEC virulence, they are not present in all AIEC strains and cannot be used to define the pathovar. The aim of this study was to compare human-derived AIEC and non-AIEC strains with similar genetic backgrounds (B2 phylogroup and ST95 lineage), isolated from the same site in the intestine, using whole-genome sequencing and other methods to identify a molecular marker of the AIEC phenotype.

MATERIALS AND METHODS

Bacterial isolates

All patients and controls attended Canberra Hospital, Australia, and gave their informed consent. Ethics approval was obtained from the hospital and university ethics committees. We preferentially selected genetically similar *E. coli* strains by focusing on the B2 phylogroup, ST95 strains, isolated from a single host species (human) and gut region (terminal ileum). This was done to eliminate among-strain variation, and to account for the possibility that AIEC has host-species or gut-region preferences.

Strain LF82, isolated from a patient with ileal CD, is the archetypal AIEC strain,[32] belongs to the B2 phylogroup of *E. coli* and was used as a positive control. Strain K12 does not display the AIEC phenotype and was used in the in vitro assays only, as a negative control. Strain ED1a was included because it was isolated from a healthy control, belongs to the B2 phylogroup and was found to be avirulent in a mouse lethality model.[33] The 41 strains in this study were isolated from 36 Australian patients with and without IBD (14 CD; 5 UC; 21 non-IBD), as described in Gordon *et al*;[34] O'Brien *et al*;[7] and Gordon *et al*.[35] Strain characteristics can be found in online supplementary material. All strains and tissues were stored at −80°C in luria broth (LB) glycerol until required. The serotype of each strain was determined using the Center for Genomic Epidemiology's online SeroType Finder tool (http://www.genomicepidemiology.org).[36] The ST was determined using the MLST methods of Wirth *et al*.[23] The *E. coli* phylogroup was determined for each strain using a quadruplex PCR described by Clermont *et al*,[18] which assigns B2 strains (and other phylogroups) based on the presence/absence of four genes (*chuA*, *yjaA*, *TspE4.C2* and *ArpA)*.

Phenotypical assays

We assessed the ability of all strains to adhere to and invade intestinal epithelial cell lines, as well as survive and replicate within macrophage cell lines, by conducting gentamicin protection assays with intestine-407 epithelial cells (American Type Culture Collection (ATCC) chemokine [c-c motif] ligand 6 (CCL-6)) and THP-1 macrophages (ATCC TIB-202), respectively, as previously described. We followed the methods of Darfeuille-Michaud *et al*,[9] except that we used the human-derived THP-1 cell line instead of the murine J774 cell line. Assays were performed in 24-well tissue culture plates, in triplicate. The AIEC strain LF82, and non-AIEC *E. coli* strain, K12, were used as controls.

To be considered AIEC, strains were required to adhere to undifferentiated I-407 epithelial cells with an adhesion index of one or more bacteria per cell; invade intestine-407 cells with an invasion index greater than 0.1% of the original inoculum; and survive and replicate within THP-1 macrophages with a survival index of 100% or greater at 24 h relative to the number of intracellular bacteria at 1 h post infection.[9]

ELISA

The amount of TNF-α, interleukin (IL) 6, IL-8 and IL-10 released into the THP-1 cell culture supernatant was determined by ELISA (R&D Systems). Cytokine concentrations were assessed according to the manufacturer's instructions. All experiments were done in triplicate.

Unsupervised iterative clustering

An unsupervised iterative clustering analysis (JMP V11, SAS Institute) was performed to determine whether or not AIEC strains clustered. This analysis does not use a priori knowledge of the AIEC status of strains, but uses the raw values from the phenotypical assays (adherence/invasion and intracellular replication) to group the isolates.

Genome sequencing and analysis

Genomic DNA was extracted from LB culture broths using Qiagen genomic kits, and quantified using a Qubit fluorescence assay (Invitrogen). All 41 strains were subjected to whole-genome sequencing using either an Illumina HiSeq 2000 platform in a 100 bp paired-end format, a Roche GS FLX 454 sequencer or a MiSeq platform (see online supplementary material). Assembly of the sequences was done in CLC Workbench using global mapping, and the draft genomes aligned in Mauve[37] and annotated in GenoScope (http://www.genoscope.cns.fr). Clustering of proteins into groups of homologous sequences was done using CD-HIT with a cut-off of 80% amino acid identity. A binary matrix was created to indicate the presence of protein clusters within strains. This matrix was statistically compared using the R (V.3.2.0) data analysis software (https://www.r-project.org), to determine the number of genes per strain; to plot histograms based on gene counts; to determine whether or not there were genes unique to/more frequent in AIEC strains compared with non-AIEC strains; and to identify genes unique to strains with high levels of invasiveness or replication that did not meet the AIEC criteria. The MicroScope Gene Phyloprofile tool[38] within Genoscope (http://www.genoscope.cns.fr) was used to confirm the findings of the latter analyses involving unique genes/gene frequencies using the following homology constraints: minimal alignment coverage of 0.8; sequence identity of 30%; and, bidirectional best hit. We used MicroScope to determine the presence/absence of genes identified in previous studies as being important for phenotypical characteristics of LF82 (*rpoE*,[39] *htrA*,[40] *dsbA*,[41] *hfq*,[42] *fimH*,[28] *lpfA*/*gipA*[43–45]) or AIEC/IBD-associated *E. coli* (*vgrG*, *hcp*, *vasD*, *vasG*, *impL*, *impK* (Type VI secretion system genes), *vat*, *insA*, *insB*,[30] *afaC*,[45] and *clb* (pks island),[46] *pduC*,[29] *K1*, *kpsmTII*, *fyuA*, *ibeA*.[17] The nucleotide sequence of each of these genes was also used for phylogenetic analyses. MEGA V.6[47] was used to align the sequences and generate phylogenetic trees.

Determining base composition differences in genes

Harvest software suite[48] was used to build a core-genome single nucleotide polymorphism-based tree from the assembled genomes. From the Mauve alignment file, nucleotide base frequency differences between AIEC and non-AIEC strains in the core genome were examined over 300 bp windows and quantified using the G-statistic. The base frequencies within each window were compared between AIEC and non-AIEC strains, as well as AIEC-ST95 strains and non-AIEC-ST95 strains. Gene sequences with meaningful different base compositions, that is, G-statistic values well above 'background' levels in AIEC compared with non-AIEC strains, and AIEC-ST95 strains compared with non-AIEC-ST95 strains (see online supplementary material), were extracted from MicroScope using the 'Search/Export' and 'Search by Keywords' functions. The gene sequences were saved in fasta file format and imported into MEGA V.6.[47] Phylogenetic trees were constructed to determine whether or not the AIEC or AIEC-ST95 gene sequences clustered together when compared with non-AIEC and non-AIEC-ST95 gene sequences, respectively.

Clustered regularly interspaced palindromic repeat analysis

Bacteria insert short sequences ('spacers'), which they acquire from invading viruses, into clustered regularly interspaced palindromic repeat (CRISPR) loci to generate immunological memory.[49] We used the CRISPRFinder web tool[50] (http://crispr.u-psud.fr/Server/) to identify CRISPRs in AIEC-ST95 and non-AIEC-ST95 strains, to determine whether or not particular CRISPRs were associated with the AIEC phenotype.

RESULTS

Strain characteristics and phenotypical assays

Of the 41 B2 isolates, 11 met the criteria for AIEC, including the AIEC strain, LF82 (figure 1). Table 1 outlines the strain characteristics, including: serotype, ST, ability to adhere to and to invade epithelial cells, ability to survive and replicate within macrophages, and affiliation to the AIEC pathovar based on phenotypical tests. Further information on strains, including raw values for all three phenotyping replicates and average $\log^{10}$ values for the adhesion, invasion and survival/replication assays, are provided in the online supplementary material.

The invasion level of the reference strain LF82 was 15.87% ±3.3%, for the laboratory K12 strain (laboratory strain) it was 0.06%±0.00% and for the commensal strain ED1a it was 0.05%±0.01% of the original inoculum (table 1). For the remaining 39 B2 isolates, 13 (33%) were considered invasive (isolated from patients with CD (n=3), patients with UC (n=3) and patients without IBD (n=7)), with invasion levels ranging from 0.10% to 1.89% (table 1).

As previously demonstrated,[51] strain ED1a was killed by THP-1 macrophages, demonstrating efficient bactericidal activity of the cell lines (table 1). As expected, strain K12 was also killed, and strain LF82 resisted macrophage killing (table 1). Of the strains 31/39 (79%) were able to resist macrophage killing and replicate within macrophages, but only 10 of these strains (25%) were also invasive, and therefore AIEC (table 1).

Proinflammatory cytokine production (TNF-α, IL-6, IL-8 and IL-10) by infected THP-1 macrophages did not differ with the

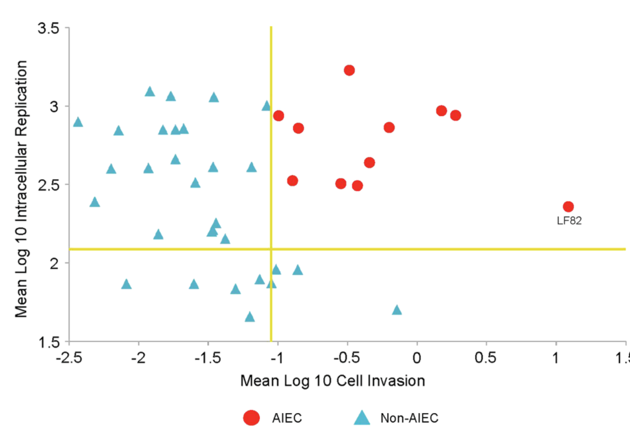

Figure 1 A plot of invasion ($\log^{10}$ bacterial cells) on the x axis, versus intracellular replication ($\log^{10}$ bacterial cells) on the y axis, showing that the majority of strains in the study can replicate intracellularly in vitro, at the level required to be considered adherent-invasive *Escherichia coli* (AIEC) (horizontal yellow bar). Fewer strains meet the AIEC invasion index (vertical yellow bar), but cannot replicate intracellularly. To be considered AIEC, strains are required to *both* invade and replicate intracellularly at the defined levels. AIEC strains are denoted by red dots, non-AIEC strains by blue triangles. I-407 intestinal cell line was used for the adherence/invasion assays, THP-1 macrophages for the survival/replication assays.

Table 1 Characteristics and adherent-invasive *Escherichia coli* (AIEC) phenotype data for the B2 phylogroup *E. coli* strains used in the study

Strain	Serotype	Sequence type (ST)*	Phenotype†	Adhesion‡	Invasion§	Replication¶
12-1-TI12	O6: H31	127	AIEC	3.81±2.03	0.37±0.11	310.60±53.48
12-2-TI13	O6: H31	127	Non-AIEC	5.61±2.31	0.72±0.13	50.24±24.33
18-3 TI5	O25:H1	73	Non AIEC	0.25±0.10	0.04±0.01	179.30±76.61
18-4 TI12	O25:H1	73	Non AIEC	0.48±0.26	0.04±0.02	143.09±25.45
33-1-TI5	O16:H5	131	AIEC	6.68±2.28	0.10±0.03	867.09±77.11
36-1-TI13	O83:H4	429	Non AIEC	0.39±0.17	0.01±0.00	402.45±86.68
39-2-TI18	** : **	95	Non AIEC	0.07±0.03	0.01±0.00	1241.09±597.43
41-2-TI13	O17/O77:H41	720	Non AIEC	0.68±0.20	0.03±0.01	410.29±85.78
46-1-TI2	O1:H7	95	Non AIEC	0.20±0.06	0.01±0.01	707.98±358.64
52-1-TI3	O18:H7	95	AIEC	34.26±9.89	1.50±0.15	934.28±250.38
52-2-TI10	O75:H5	550	Non AIEC	0.11±0.08	0.01±0.00	153.11±31.56
54-1-TI6	O110:H27	1919	Non AIEC	9.30±5.13	0.07±0.03	408.36±65.09
55-1-AU4	O25:H4	131	Non AIEC	0.11±0.04	0.03±0.02	162.92±82.29
55-1-TI19	O25:H4	131	Non AIEC	0.05±0.05	0.03±0.01	325.74±164.91
57-3-TI5	O75:H5	537	AIEC	2.73±1.07	0.14±0.03	723.55±65.60
60-1-TI1	O75:H7	80	AIEC	7.08±1.35	0.13±0.06	333.25±107.54
61-1-TI1	O18:H7	95	AIEC	54.63±16.65	1.89±0.70	869.90±211.49
62-2-TI6	O1:H7	95	Non AIEC	1.24±0.33	0.14±0.02	90.74±35.52
63-1-TI1	O50O2:H18	963	Non AIEC	1.43±0.63	0.03±0.00	158.22±42.48
69-1 AU1	O46:H31	569	Non AIEC	0.66±0.15	0.07±0.03	79.02±12.65
69-1-TI1	O46:H31	569	Non-AIEC	0.60±0.13	0.10±0.03	91.37±30.65
70-2-TI12	O46:H31	569	Non-AIEC	1.22±0.58	0.06±0.01	45.49±5.68
72-6-Ti12	O46:H31	569	Non-AIEC	1.05±0.06	0.03±0.00	73.82±17.73
CD 1IM 3	O83:H6	2622	Non-AIEC	3.47±0.27	0.02±0.00	1163.70±203.59
CD 34 LN	O1:H7	95	Non-AIEC	0.14±0.06	0.02±0.00	458.485±105.69
CD 62 LN	O2:H4	95	AIEC	4.68±2.07	0.33±0.06	1686.33±307.68
H001	O8:H10	681	Non-AIEC	1.19±0.02	0.02±0.00	708.12±55.32
H020	O22:H1	73	Non-AIEC	1.33±0.07	0.08±0.04	1009.20±137.15
H223	O2:H6	141	AIEC	9.03±2.55	0.46±0.23	435.23±108.95
H252	O1:H7	95	Non-AIEC	3.47±0.20	0.09±0.04	74.23±30.72
H263	O2:H7	95	Non-AIEC	0.30±0.03	0.01±0.00	398.51±93.69
H296	O2:H7	95	Non-AIEC	0.45±0.04	0.00±0.00	792.33±78.74
H305	O18:H7	95	AIEC	27.52±6.10	0.63±0.24	730.65±68.44
H397	O1:H7	95	Non-AIEC	0.42±0.09	0.01±0.00	245.22±50.18
H413	O18:H7	95	Non-AIEC	0.91±0.05	0.02±0.00	717.92±411.86
H461	O18:H7	95	Non-AIEC	0.40±0.07	0.01±0.00	73.60±19.38
H504	O18ac:H7	95	AIEC	3.28±1.78	0.28±0.03	330.81±80.68
H588	** :H5	126	Non-AIEC	3.38±1.39	0.04±0.00	1141.51±65.45
H660	O1:H7	95	Non-AIEC	0.52±0.16	0.01±0.00	701.56±145.06
ED1a	O81:H27	452	Non-AIEC	0.25±0.07	0.05±0.01	68.34±30.95
LF82	O83:H1	135	AIEC	62.83±5.08	12.23±2.01	227.81±61.37
K-12	OR:H48	10	Non-AIEC	1.43±0.40	0.02±0.00	22.49±4.34

Results for adhesion, invasion and replication assays, data are means±SEM of independent experiments.
*Sequence type (ST) is based on the multilocus sequence type (MLST) scheme of Wirth *et al.*[23]
†Determination of *E. coli* strains as belonging to the AIEC pathovar was performed using the following criteria: (1) the ability to adhere to I-407 epithelial cells with an adhesion index equal or superior to 1 bacteria per cell, (2) the ability of the bacteria to invade I-407 with an invasion index equal or superior to 0.1% of the original inoculum, (3) the ability to survive and to replicate within THP-1 macrophages.
‡Mean number of bacteria per I-407 cell after 3 h of incubation.
§Mean percentage of the original inoculum after 1 h gentamicin treatment of infected I-407 cells.
¶Percentage of intracellular bacteria at 24 h post infection relative to the number after 1 h gentamicin treatment, defined as 100%.
**No O-antigen processing genes and/or H-antigen flagellin genes detected using the SeroType Finder tool (http://www.genomicepidemiology.org).
LN, lymph node.

AIEC phenotype of a strain (analysis of variance: TNF-α, P> F=0.88; IL-6, P> F=0.99). IL-10 production correlated positively with the invasiveness of the strain.

Unsupervised clustering analysis confirms AIEC phenotype

An unsupervised clustering analysis, based on the raw values from the adhesion/invasion assays, performed using I-407 epithelial cells, and survival/replication assays, using THP-1 macrophages, shows that AIEC and non-AIEC strains cluster in two distinct groups (figure 2), suggesting that these functions commonly coexist in a specific isolate. Strain LF82 is more similar to the AIEC strains than non-AIEC strains, but is an outlier. The one non-AIEC strain in the AIEC cluster (strain H020) demonstrated high levels of intracellular replication but was on the borderline for cellular invasion (0.08%).

Phylogenetic distribution and serotyping

The genomes in this study were a typical size for *E. coli*, ranging from 4.4 Mb to 5.4 Mb, and had an average 56×read depth (see online supplementary material). The core-genome phylogenetic

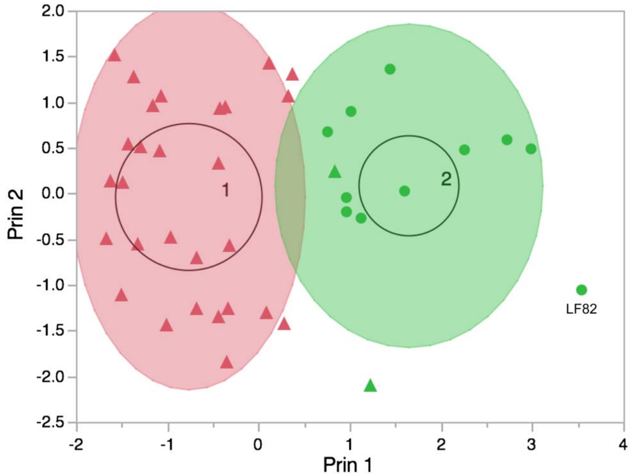

Figure 2 Unsupervised iterative clustering analysis showing that adherent-invasive *Escherichia coli* (AIEC) strains (dots), and non-AIEC strains (triangles) naturally group together based on the log10 values from the adherence/invasion (I-407 cell line), and survival/replication (THP-1 macrophage cell line) assays. Strain LF82, shown at the right of the graph, is more similar to AIEC strains than non-AIEC strains, but is an outlier. The axes represent Principal 1, Prin 1, and Principal 2, Prin 2, of the analysis.

tree presented in figure 3 shows that the 11 AIEC isolates have diverse B2 phylogenetic backgrounds, as they are dispersed among the 31 non-AIEC strains, and represented across seven B2 lineages (ST131, ST127, ST80, ST537, ST135, ST141, ST95). This is despite strain selection being biased towards the ST95 lineage. The AIEC strains also represent diverse serotypes (O16:H5, O6:H31, O75:H7, O75:H5, O83:H1, O2:H6, O2:H4, O2:H7, O18:H7 and O18ac:H7). Of the ST95 strains, four of six strains with the O18:H7 serotype are AIEC (table 1).

Comparative genome analyses

The AIEC strains did not harbour any unique genes when compared with the non-AIEC strains, nor were there any gene(s) present in the majority of AIEC strains, but absent in non-AIEC strains. There were no genes present in all non-AIEC strains that were absent from all AIEC strains. There were no genes unique to strains with a replication index (using THP-1 macrophage cell line) that met the AIEC criteria, but a level of adhesion/invasion (using I-407 cell line) that did not; strains with an adherence/invasion index that met the criteria for AIEC, but a level of replication that did not; strains with a high capability to replicate within macrophages, having a replication index of >700, or >1000 (where the percentage of the number of intracellular bacteria at 24 h post infection relative to that obtained at 1 h post infection is defined as 100%), regardless of their adhesion/invasion index; or strains from patients with CD compared with controls.

Two of the 41 strains, 12–1 ti12 and 12–2 ti13, were isolated from the same patient but displayed different AIEC phenotypes. The genomes of these two strains differed by 151 genes, however none of the genes unique to the AIEC strain, 12–1 ti12, were present in all, or the majority of, the other AIEC strains, and none of the genes unique to strain 12–2 ti13 were absent from all, or the majority of, other AIEC strains. The variable gene content of these two strains is presented in the online supplementary material.

We conducted presence/absence and phylogenetic analyses on the following LF82-associated genes: *rpoE*;[39] *htrA* (also known as *degP* and *ptd*);[40] *dsbA*,[41] *hfq*,[42] and AIEC/IBD-associated *E. coli genes*: *pduC*[29] *(propanediol utilisation)*, *fyuA*, *ibeA*, *K1*, *kpsmTII*,[17] *vgrG*, *hcp*, *vasD*, *vasG*, *impL*, *impK* (type VI secretion system genes), and *vat*, *insA*, *insB*,[30] *fimH*,[28] *afaC*[45] and *clb* (pks island),[46] and the combination of *gipA* and *lpfA*,[43] or *lpfA* alone.[44 45] Nine of the 41 strains harboured the *afaC* gene (H504, H305, H020, H296, 61–1 ti1, 52–2 ti10, 52–1 ti3, 18–4 ti12, 18–3 ti5), 4 of which were AIEC (36% of all AIEC strains), belonged to ST95 and had an O18:H7 serotype. The only gene that was detected in all AIEC strains was *fimH*, however this gene is present in most B2 phylogroup *E. coli* strains. Differences in the base composition of all the above-mentioned genes were largely driven by ST, thereby phylogenetic in nature, not the AIEC status of the strains. The most common finding was that ST95 strains had a different variant (base composition) of a gene than other STs. One exception to this was *fimH*, as the *fimH* variant of ST95 strains with an O18:H7 serotype was different to that of other ST95 strains with a different serotype.

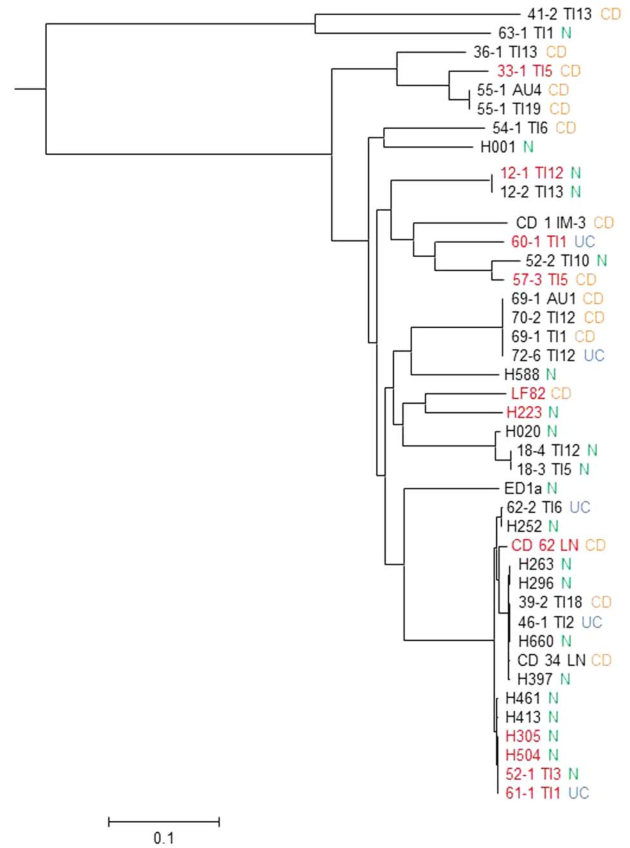

Figure 3 Core-genome phylogenetic tree for all 41 of the B2 phylogroup strains of *Escherichia coli* used in the study, showing the position of the 11 adherent-invasive *E. coli* (AIEC) strains (red). The AIEC strains have diverse genetic backgrounds and are spread throughout the tree. An orange 'Crohn's disease (CD)' indicates strains isolated from patients with CD, a blue 'UC' strain from patients with UC and a green 'N' strain from patients without IBD. The sequence type of each strain is indicated on the right, and follows the multilocus sequence type (MLST) classification scheme by Wirth *et al*. AIEC strains display phylogenetic heterogeneity, as they are present across a large number of STs within the B2 phylogenetic lineage.

Differences in base composition of genes

We identified numerous genes where the base composition in AIEC strains differed from that in non-AIEC strains over a

300 bp window. We conducted phylogenetic analyses on the following genes, which met inclusion criteria (G-statistic >0.05, MI> 0.01, see online supplementary material): *nagE*, *ompN*, *ompF*, *ugd*, *dusA*, *yghJ*, *fkpA*, *btuB*, *prp* and *yehH*. When the AIEC-ST95 and non-AIEC-ST95 strains were compared, the following genes were identified: *yjjM*, *yjjN*, *galF*, *gnd*, *hchA* and *wcaK,* and analysed phylogenetically. For each of these genes, either there was no clustering of the AIEC strains, or strains clustered according to lineage (based on the entire core genome), irrespective of their AIEC status. The online supplementary material outlines all genes shown to have meaningfully different base frequencies when AIEC and non-AIEC, and, AIEC-ST95 and non-AIEC-ST95 strains were compared.

CRISPR analyses

Among the 16 ST95 strains tested, we identified 23 confirmed CRISPRs and 26 possible CRISPRs, with an average of 2.4 confirmed CRISPRs (range: 2–5) and 3.3 possible CRISPRs (range: 0–8) per strain. AIEC-ST95 strains did not harbour specific CRISPRs, nor was the frequency of CRISPRs, or possible CRISPRs, significantly different to that of non-AIEC-ST95 strains.

CRISPR analysis of closely related strains 12–1 (AIEC) and 12–2 ti13 (non-AIEC), from the same patient, and 55–1 AU4 and 55–1 ti19 from another patient, demonstrated different CRISPR profiles. Strain 12–1 ti12 possessed four possible CRISPRs, whereas strain 12–2 ti13 possessed five possible CRISPRs, none of which were found in strain 12–1 ti12 (see online supplementary material). Strain 55–1 AU4 and 55–1 ti19 both had three possible CRISPRs, they shared two of these and had one unique CRISPR each.

DISCUSSION

One of the difficulties associated with *E. coli* comparative genomics is the among-strain variability in gene content.[33] An *E. coli* genome typically consists of about 4500 genes, but less than half of these are core genes present in all strains. The balance of genes is drawn from a pool of more than 14 000 unique accessory genes. In an attempt to minimise the among-strain differences in gene content, we chose to work preferentially with strains belonging to a single phylogroup (B2), and to further refine our focus, strains that were ST95. We selected strains that were isolated from a single gut region, to account for any possible niche differences. Despite this choice of strains, we did not identify a single or multiple genes of the variable genome associated with all, or the majority of, our Australian AIEC isolates.

AIEC employ different sets or variants of genes, to overcome mechanical forces in the gut, mucosal defenses, and subvert antimicrobial macrophage pathways. We show that no single gene is associated with the ability of the strains to invade epithelial cells, or to survive/replicate within macrophages. UPEC also overcome numerous defences in order to adhere to and invade the urinary tract, and as we have shown for AIEC, there is no specific gene associated with UPEC. While some factors, such as type 1 pili are important for UPEC to be able to establish an infection, these factors are not restricted to UPEC. *E. coli* have a range of mechanisms that enable them to invade epithelial cells and replicate in macrophages; these factors are not exclusive to AIEC. Given the multiple processes facilitating invasion and replication, and source of origin (GI lumen) of AIEC and UPEC, it is likely that these two pathovars overlap.

The Intestine-407 (I-407) cell line was used in this study to evaluate adhesion and invasion abilities of *E. coli* strains to epithelial cells, as in the original description of the AIEC pathovar.[9 15] This cell line actually resulted from hela cell (HeLa) contamination and is of cervical carcinoma origin, not embryonic intestinal origin as originally thought,[52] but is still used as a model for measuring the ability of intestinal bacteria to adhere to and to invade epithelial cells.[45 53] Other studies of AIEC used cancerous cell lines such as Caco-2[16] and HT-29 cell lines,[11] derived from colorectal carcinomas, and the HEp-2 cell line,[17] which is also the result of HeLa contamination. It is not clear how applicable these cell lines are to CD pathogenesis, given their non-intestinal, transformed or cancerous origin.

It is plausible, especially given the lack of empirical evidence demonstrating the presence of AIEC within human epithelial cells, that AIEC use other routes to invade intestinal mucosa, for example by translocation across M cells lining the follicle-associated epithelia. Many other pathogens exploit this route of entry, including *enteropathogenic E. coli* and *enterohaemorrhagic E. coli* to colonise the intestinal epithelium. Chassaing *et al*[54] showed that some AIEC interact with human and mouse Peyer's patches through the expression of type 1 pili and LPF. AIEC type 1 pili-mediated interaction with CEACAM6 could also disrupt barrier integrity giving bacteria access to the subepithelial compartment.[55] In addition, the ability to adhere to and invade intestinal epithelial cells may not be required in the presence of mucosal ulceration, and may not be relevant to the pathogenesis of CD.

The ability of *E. coli* to adhere to and invade different cell lines depends both on bacterial factors, such as the expression of various adhesion or invasion factors, and the specific cell line used. Martin *et al*[11] showed that colonic mucosa-associated *E. coli* strains from CD and colon cancer differed in their ability to invade two epithelial cell lines: I-407 and HT29. The HT29 cell line is of human colon adenocarcinoma origin, and the vast majority of strains adhered poorly to these cells compared with I-407 cells: 74% of strains invaded I-407 cells, but only 9% of strains invaded HT29 cells. AIEC strain LF82 was a notable exception, invading both cell lines at a similar rate. Boudeau *et al*[15] found that strains LF82 and K12 displayed consistent levels of invasion across three different cell lines (I-407, Caco-2 and HCT-8), but that enteroinvasive *E. coli* reference strain E12860/0, which invades colonic epithelial cells in vivo, showed inconsistent levels of invasion. They invaded Caco-2 cells at a very low level, I-407 cells at an intermediate level similar to LF82 and HCT-8 cells at a much higher rate than any strain tested, including LF82.

There are also strain-dependent differences in intramacrophage survival and replication of *E. coli*. Bokil *et al*[56] found that clinical UPEC isolates performed differently in murine and human macrophages. We used the human THP-1 myelomonocytic cell line that displays macrophage-like activity. It is not known if the more widely used murine J774 reticulosarcoma cell line results in the same AIEC designation. A cross-validation study is underway to determine whether or not different cell lines and scoring criteria result in the same AIEC designation; the outcome may result in a standardised approach to AIEC phenotyping. These studies may help in terms of defining the phenotype, AIEC, however we cannot be certain that the in vitro behaviour of strains reflects their in vivo behaviour.

We found that 79% of our B2 phylogroup *E. coli* strains, irrespective of AIEC status, could survive and replicate within macrophages. Subramanian *et al*[57] also observed considerable overlap between the ability of CD and control isolates to replicate within macrophages, and Raisch *et al*[51] showed that 84% of colon cancer-associated B2 *E. coli* strains can survive and replicate within THP-1 macrophages. UPEC are also capable of intramacrophage replication.[56] A recent study showed that *E.*

coli strains isolated from patients with CD, irrespective of their adherent, invasive phenotype, survived longer in monocyte-derived macrophages isolated from patients with CD than controls, demonstrating that host immunodeficiency is an important factor allowing *E. coli* to persist.[58] These findings implicate a broader set of *E. coli* in CD pathogenesis: those that are capable of intramacrophage survival and replication. The use of macrophages isolated from patients with CD, and the isolation of strains from more relevant tissues, such as aphthous ulcers and lymph nodes, may represent a better model to study host-microbe interactions in CD. In such experiments, host genetic susceptibility factors, and particularly those related to detection and intracellular control of invaders (polymorphisms in pattern recognition receptor (PRRs) and autophagy-related genes) should be taken into account, as they will influence the intracellular behaviour of bacteria.[59]

One of our AIEC strains, isolated from a lymph node of a CD bowel resection, showed the highest level of replication of all strains. Another non-AIEC strain, isolated from a lymph node of a different patient with CD, was better at replicating intracellularly than LF82. Intracellular replication may be more important than a strain's ability to adhere/invade epithelial cells, because host mutations in genes involved in macrophage function likely lead to increased intramacrophage survival. It is plausible that defective macrophages serve as bacterial reservoirs, like the quiescent intracellular reservoirs and intracellular bacterial communities characteristic of UPEC.

In conclusion, we were unable to identify a specific molecular property of the AIEC phenotype by comparing genomes, gene variants and base composition differences of genetically similar AIEC and non-AIEC strains. Studying the interactions of a broader range of *E. coli* and their interactions, for example, with monocyte-derived macrophages isolated from patients with CD, may provide further insights into additional or alternative pathogenic mechanisms.

Acknowledgements This manuscript is dedicated to the memory of AD-M. The authors thank the subjects who provided the specimens for this research.

Contributors Study concept and design (CLOB, M-AB, DMG and PP); acquisition of data (CLOB, M-AB, KEH, DMG and PP); analysis and interpretation of data (CLOB, M-AB, KEH, DMG and PP); drafting of the manuscript (CLOB); critical revision of the manuscript for important intellectual content (CLOB, M-AB, KEH, DMG, NB and PP); statistical analysis (CLOB, M-AB, KEH and DMG); obtained funding (CLOB, M-AB, AD-M and PP); technical or material support (CLOB, M-AB, KEH, DMG, ALD, NB, AD-M and PP).

Funding This research was supported by an Australian Academy of Science France-Australia Science Innovation Collaboration early career fellowship; a Gastroenterological Society of Australia (GESA) Clinical Research grant; and funding from the Ministère de la Recherche et de la Technologie, Inserm (UMR1071), INRA (USC-2018); and Nouveau Chercheur EPST from Conseil Régional Auvergne. Study sponsors did not influence or dictate the study design, analysis or interpretation of the data.

Competing interests None declared.

Patient consent Obtained.

Ethics approval ACT Health Human Research Ethics Committee, and, Australian National University Human Ethics Research Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All GenBank files (genome data), and associated metadata are available to the public, and can be accessed via the following doi: (10.4225/13/56F08C5B5F0FE). Please cite this publication.

REFERENCES

1 Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* 2014;146:1489–99.
2 Jostins L, Ripke S, Weersma RK, *et al*. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;491:119–24.
3 Liu JZ, van Sommeren S, Huang H, *et al*. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* 2015;47:979–86.
4 Gevers D, Kugathasan S, Denson LA, *et al*. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 2014;15:382–92.
5 Kang S, Denman SE, Morrison M, *et al*. Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm Bowel Dis* 2010;16:2034–42.
6 Morgan XC, Tickle TL, Sokol H, *et al*. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 2012;13:R79.
7 O'Brien CL, Pavli P, Gordon DM, *et al*. Detection of bacterial DNA in lymph nodes of Crohn's disease patients using high throughput sequencing. *Gut* 2014;63:1596–606.
8 Frank DN, St Amand AL, Feldman RA, *et al*. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci USA* 2007;104:13780–5.
9 Darfeuille-Michaud A, Boudeau J, Bulois P, *et al*. High prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in Crohn's disease. *Gastroenterology* 2004;127:412–21.
10 Kotlowski R, Bernstein CN, Sepehri S, *et al*. High prevalence of Escherichia coli belonging to the B2+D phylogenetic group in inflammatory bowel disease. *Gut* 2007;56:669–75.
11 Martin HM, Campbell BJ, Hart CA, *et al*. Enhanced Escherichia coli adherence and invasion in Crohn's disease and colon cancer. *Gastroenterology* 2004;127:80–93.
12 Martinez-Medina M, Aldeguer X, Lopez-Siles M, *et al*. Molecular diversity of Escherichia coli in the human gut: new ecological evidence supporting the role of adherent-invasive E. coli (AIEC) in Crohn's disease. *Inflamm Bowel Dis* 2009;15:872–82.
13 Mylonaki M, Rayment NB, Rampton DS, *et al*. Molecular characterization of rectal mucosa-associated bacterial flora in inflammatory bowel disease. *Inflamm Bowel Dis* 2005;11:481–7.
14 Sasaki M, Sitaraman SV, Babbin BA, *et al*. Invasive Escherichia coli are a feature of Crohn's disease. *Lab Invest* 2007;87:1042–54.
15 Boudeau J, Glasser AL, Masseret E, *et al*. Invasive ability of an Escherichia coli strain isolated from the ileal mucosa of a patient with Crohn's disease. *Infect Immun* 1999;67:4499–509.
16 Simpson KW, Dogan B, Rishniw M, *et al*. Adherent and invasive Escherichia coli is associated with granulomatous colitis in boxer dogs. *Infect Immun* 2006;74:4778–92.
17 Conte MP, Longhi C, Marazzato M, *et al*. Adherent-invasive Escherichia coli (AIEC) in pediatric Crohn's disease patients: phenotypic and genetic pathogenic features. *BMC Res Notes* 2014;7:748.
18 Clermont O, Christenson JK, Denamur E, *et al*. The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* 2013;5:58–65.
19 Gordon DM, O'Brien CL, Pavli P. Escherichia coli diversity in the lower intestinal tract of humans. *Environ Microbiol Rep* 2015;7:642–8.
20 Tenaillon O, Skurnik D, Picard B, *et al*. The population genetics of commensal Escherichia coli. *Nat Rev Microbiol* 2010;8:207–17.
21 Jaureguy F, Landraud L, Passet V, *et al*. Phylogenetic and genomic diversity of human bacteremic Escherichia coli strains. *BMC Genomics* 2008;9:560.
22 Reid SD, Herbelin CJ, Bumbaugh AC, *et al*. Parallel evolution of virulence in pathogenic Escherichia coli. *Nature* 2000;406:64–7.
23 Wirth T, Falush D, Lan R, *et al*. Sex and virulence in Escherichia coli: an evolutionary perspective. *Mol Microbiol* 2006;60:1136–51.
24 Clermont O, Christenson JK, Daubié AS, *et al*. Development of an allele-specific PCR for Escherichia coli B2 sub-typing, a rapid and easy to perform substitute of multilocus sequence typing. *J Microbiol Methods* 2014;101:24–7.
25 Martinez-Medina M, Mora A, Blanco M, *et al*. Similarity and divergence among adherent-invasive Escherichia coli and extraintestinal pathogenic E. coli strains. *J Clin Microbiol* 2009;47:3968–79.
26 Barnich N, Carvalho FA, Glasser AL, *et al*. CEACAM6 acts as a receptor for adherent-invasive E. coli, supporting ileal mucosa colonization in Crohn disease. *J Clin Invest* 2007;117:1566–74.
27 Carvalho FA, Barnich N, Sivignon A, *et al*. Crohn's disease adherent-invasive Escherichia coli colonize and induce strong gut inflammation in transgenic mice expressing human CEACAM. *J Exp Med* 2009;206:2179–89.
28 Dreux N, Denizot J, Martinez-Medina M, *et al*. Point mutations in FimH adhesin of Crohn's disease-associated adherent-invasive Escherichia coli enhance intestinal inflammatory response. *PLoS Pathog* 2013;9:e1003141.
29 Dogan B, Suzuki H, Herlekar D, *et al*. Inflammation-associated adherent-invasive Escherichia coli are enriched in pathways for use of propanediol and iron and M-cell translocation. *Inflamm Bowel Dis* 2014;20:1919–32.

30 Desilets M, Deng X, Rao C, *et al*. Genome-based definition of an inflammatory bowel disease-associated adherent-invasive Escherichia coli pathovar. *Inflamm Bowel Dis* 2016;22:1–12.
31 Deshpande NP, Wilkins MR, Mitchell HM, *et al*. Novel genetic markers define a subgroup of pathogenic Escherichia coli strains belonging to the B2 phylogenetic group. *FEMS Microbiol Lett* 2015;362.
32 Miquel S, Peyretaillade E, Claret L, *et al*. Complete genome sequence of Crohn's disease-associated adherent-invasive E. coli strain LF82. *PLoS ONE* 2010;5.
33 Touchon M, Hoede C, Tenaillon O, *et al*. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. *PLoS Genet* 2009;5:e1000344.
34 Gordon DM, O'Brien CL, Pavli P. Escherichia coli diversity in the lower intestinal tract of humans. *Environ Microbiol Rep* 2015;7:642–8.
35 Gordon DM, Stern SE, Collignon PJ. Influence of the age and sex of human hosts on the distribution of Escherichia coli ECOR groups and virulence traits. *Microbiology* 2005;151(Pt 1):15–23.
36 Joensen KG, Tetzschner AM, Iguchi A, *et al*. Rapid and easy in silico serotyping of Escherichia coli isolates by use of whole-genome sequencing data. *J Clin Microbiol* 2015;53:2410–26.
37 Darling AC, Mau B, Blattner FR, *et al*. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004;14:1394–403.
38 Vallenet D, Belda E, Calteau A, *et al*. MicroScope--an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res* 2013;41(Database issue):D636–47.
39 Chassaing B, Garénaux E, Carriere J, *et al*. Analysis of the σE regulon in Crohn's disease-associated Escherichia coli revealed involvement of the waaWVL operon in biofilm formation. *J Bacteriol* 2015;197:1451–65.
40 Bringer MA, Barnich N, Glasser AL, *et al*. HtrA stress protein is involved in intramacrophagic replication of adherent and invasive Escherichia coli strain LF82 isolated from a patient with Crohn's disease. *Infect Immun* 2005;73:712–21.
41 Bringer MA, Rolhion N, Glasser AL, *et al*. The oxidoreductase DsbA plays a key role in the ability of the Crohn's disease-associated adherent-invasive Escherichia coli strain LF82 to resist macrophage killing. *J Bacteriol* 2007;189:4860–71.
42 Simonsen KT, Nielsen G, Bjerrum JV, *et al*. A role for the RNA chaperone Hfq in controlling adherent-invasive Escherichia coli colonization and virulence. *PloS ONE* 2011;6:e16387.
43 Vazeille E, Chassaing B, Buisson A, *et al*. GipA factor supports colonization of Peyer's Patches by Crohn's disease-associated Escherichia coli. *Inflamm Bowel Dis* 2016;22:68–81.
44 Chassaing B, Darfeuille-Michaud A. [The interaction of Crohn's disease-associated Escherichia coli to Peyer's patches of the intestinal mucosa involves long polar fimbriae]. *Med Sci (Paris)* 2011;27:572–3.
45 Prorok-Hamon M, Friswell MK, Alswied A, *et al*. Colonic mucosa-associated diffusely adherent afaC+ Escherichia coli expressing lpfA and pks are increased in inflammatory bowel disease and colon cancer. *Gut* 2014;63:761–70.
46 Arthur JC, Perez-Chanona E, Mühlbauer M, *et al*. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* 2012;338:120–3.
47 Tamura K, Stecher G, Peterson D, *et al*. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 2013;30:2725–9.
48 Treangen TJ, Ondov BD, Koren S, *et al*. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* 2014;15:524.
49 Nuñez JK, Lee AS, Engelman A, *et al*. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 2015;519:193–8.
50 Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 2007;35:W52–7.
51 Raisch J, Rolhion N, Dubois A, *et al*. Intracellular colon cancer-associated Escherichia coli promote protumoral activities of human macrophages by inducing sustained COX-2 expression. *Lab Invest* 2015;95:296–307.
52 Masters JR. HeLa cells 50 years on: the good, the bad and the ugly. *Nat Rev Cancer* 2002;2:315–19.
53 Nishiyama K, Seto Y, Yoshioka K, *et al*. Lactobacillus gasseri SBT2055 reduces infection by and colonization of Campylobacter jejuni. *PloS ONE* 2014;9:e108827.
54 Chassaing B, Rolhion N, de Vallée A, *et al*. Crohn disease--associated adherent-invasive E. coli bacteria target mouse and human Peyer's patches via long polar fimbriae. *J Clin Invest* 2011;121:966–75.
55 Denizot J, Sivignon A, Barreau F, *et al*. Adherent-invasive Escherichia coli induce claudin-2 expression and barrier defect in CEABAC10 mice and Crohn's disease patients. *Inflamm Bowel Dis* 2012;18:294–304.
56 Bokil NJ, Totsika M, Carey AJ, *et al*. Intramacrophage survival of uropathogenic Escherichia coli: differences between diverse clinical isolates and between mouse and human macrophages. *Immunobiology* 2011;216:1164–71.
57 Subramanian S, Roberts CL, Hart CA, *et al*. Replication of colonic Crohn's disease mucosal Escherichia coli isolates within macrophages and their susceptibility to antibiotics. *Antimicrob Agents Chemother* 2008;52:427–34.
58 Elliott TR, Hudspith BN, Rayment NB, *et al*. Defective macrophage handling of Escherichia coli in Crohn's disease. *J Gastroenterol Hepatol* 2015;30:1265–74.
59 Nguyen HT, Lapaquette P, Bringer MA, *et al*. Autophagy and Crohn's disease. *J Innate Immun* 2013;5:434–43.

GUT

Comparative genomics of Crohn's disease-associated adherent-invasive *Escherichia coli*

Claire L O'Brien, Marie-Agnès Bringer, Kathryn E Holt, David M Gordon, Anaëlle L Dubois, Nicolas Barnich, Arlette Darfeuille-Michaud and Paul Pavli

Gut published online April 20, 2016