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

To cite this version:

HAL Id: hal-01393979
https://hal-univ-bourgogne.archives-ouvertes.fr/hal-01393979
Submitted on 26 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License
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 Pavli1,2

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 exclusive to the AIEC phenotype. 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.

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. 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. 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. Studies show an increase in mucosa-associated Escherichia coli in both ileum and colon, but no single causative micro-organism has been identified.

The adherent-invasive E. coli (AIEC) pathovar was first described by Boudeau et al15 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 invade epithelial cells, and to survive and replicate within macrophages. 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. 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. The majority
(64%) of AIEC isolates belong to the B2 phylogroup, and this phylogroup typically contains more virulence factors than strains belonging to other phylogenoids. Multilocus sequence typing (MLST) is also used to assign E. coli into sequence types (STs), using one of three schemes. 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. ST95 strains are commonly isolated from humans and belong to the B2 phylogroup.

AIEC are more commonly isolated from mucosal biopsies of patients with CD (36–52%) than controls (6–17%), and one study showed that only 6.3% of extraintestinal pathogenic E. coli (ExPEC) strains were AIEC. 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. Point mutations in the FimH adhesion of AIEC result in an increased ability to adhere to CEACAM6-expressing intestinal epithelial cells. AIEC also target M cells on Peyers 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 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 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 dehydratase) 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 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 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 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 archetype AIEC strain, 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. 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; O’Brien et al; and Gordon et al. Strain characteristics can be found in online supplementary material. All strains and tissues were stored at −80°C in lura 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). The ST was determined using the MLST methods of Wirth et al. The E. coli phylogroup was determined for each strain using a quadruplex PCR described by Clermont et al, which assigns B2 strains (and other phylogenoids) 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, 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 1 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.

**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 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 (V3.2.0) 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 Phylolproteome tool 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, htrA, dslA, hfd, fimH, lpfA/gipA, pks island), or AIEC/IBD-associated E. coli (vgrG, hcp, vasD, vasG, impL, impK (Type VI secretion system genes), vat, insA, insB, alfA, alfC, and clb (pks island), pdcC, K1, kpsMTII, fyuA, ikeA. The nucleotide sequence of each of these genes was also used for phylogenetic analyses. MEGA V6.47 was used to align the sequences and generate phylogenetic trees.

Determining base composition differences in genes
Harvest software suite 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 V6.45 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 (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 log10 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).

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

---

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

<table>
<thead>
<tr>
<th>Strain</th>
<th>Serotype</th>
<th>Sequence type (ST)*</th>
<th>Phenotype†</th>
<th>Adhesion‡</th>
<th>Invasion§</th>
<th>Replication¶</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-1-TI12</td>
<td>O6: H31</td>
<td>127</td>
<td>AIEC</td>
<td>3.81±2.03</td>
<td>0.37±0.11</td>
<td>310.60±53.48</td>
</tr>
<tr>
<td>12-2-TI13</td>
<td>O6: H31</td>
<td>127</td>
<td>Non-AIEC</td>
<td>5.61±2.31</td>
<td>0.27±0.12</td>
<td>502.42±24.33</td>
</tr>
<tr>
<td>18-3-T15</td>
<td>O25:H1</td>
<td>73</td>
<td>Non AIEC</td>
<td>0.25±0.10</td>
<td>0.04±0.01</td>
<td>179.30±76.61</td>
</tr>
<tr>
<td>18-4-T12</td>
<td>O25:H1</td>
<td>73</td>
<td>Non AIEC</td>
<td>0.48±0.26</td>
<td>0.04±0.02</td>
<td>143.09±25.45</td>
</tr>
<tr>
<td>33-1-T15</td>
<td>O16:H5</td>
<td>131</td>
<td>AIEC</td>
<td>6.68±2.28</td>
<td>0.10±0.03</td>
<td>867.09±77.11</td>
</tr>
<tr>
<td>36-1-TI13</td>
<td>O83:H4</td>
<td>429</td>
<td>AIEC</td>
<td>0.39±0.17</td>
<td>0.01±0.00</td>
<td>402.45±86.68</td>
</tr>
<tr>
<td>39-2-TI18</td>
<td>* *</td>
<td>95</td>
<td>Non AIEC</td>
<td>0.07±0.03</td>
<td>0.01±0.00</td>
<td>1241.09±597.43</td>
</tr>
<tr>
<td>41-2-TI13</td>
<td>O17/077:H41</td>
<td>720</td>
<td>Non AIEC</td>
<td>0.68±0.20</td>
<td>0.03±0.01</td>
<td>410.29±85.78</td>
</tr>
<tr>
<td>46-1-TI2</td>
<td>O1:H7</td>
<td>95</td>
<td>Non AIEC</td>
<td>0.20±0.06</td>
<td>0.01±0.00</td>
<td>707.98±358.64</td>
</tr>
<tr>
<td>52-1-TI3</td>
<td>O18:H7</td>
<td>95</td>
<td>AIEC</td>
<td>34.26±9.89</td>
<td>1.50±0.15</td>
<td>934.28±250.38</td>
</tr>
<tr>
<td>52-2-T110</td>
<td>O75:H5</td>
<td>550</td>
<td>Non AIEC</td>
<td>0.11±0.08</td>
<td>0.01±0.00</td>
<td>153.11±31.56</td>
</tr>
<tr>
<td>54-1-TI6</td>
<td>O110:H27</td>
<td>1919</td>
<td>Non AIEC</td>
<td>0.48±0.26</td>
<td>0.04±0.02</td>
<td>143.09±25.45</td>
</tr>
<tr>
<td>55-1-AU4</td>
<td>O25:H4</td>
<td>131</td>
<td>Non AIEC</td>
<td>0.05±0.05</td>
<td>0.03±0.01</td>
<td>325.74±164.91</td>
</tr>
<tr>
<td>55-1-TI19</td>
<td>O25:H4</td>
<td>131</td>
<td>Non AIEC</td>
<td>0.05±0.05</td>
<td>0.03±0.01</td>
<td>325.74±164.91</td>
</tr>
<tr>
<td>57-3-T15</td>
<td>O75:H5</td>
<td>537</td>
<td>AIEC</td>
<td>2.73±1.07</td>
<td>0.14±0.03</td>
<td>723.55±65.60</td>
</tr>
<tr>
<td>60-1-TI1</td>
<td>O75:H7</td>
<td>80</td>
<td>AIEC</td>
<td>7.08±1.35</td>
<td>0.13±0.06</td>
<td>333.25±107.54</td>
</tr>
<tr>
<td>61-1-TI1</td>
<td>O18:H7</td>
<td>95</td>
<td>AIEC</td>
<td>54.63±16.65</td>
<td>1.89±0.70</td>
<td>869.90±211.49</td>
</tr>
<tr>
<td>62-2-TI6</td>
<td>O1:H7</td>
<td>95</td>
<td>Non AIEC</td>
<td>1.24±0.33</td>
<td>0.06±0.01</td>
<td>73.82±17.73</td>
</tr>
<tr>
<td>63-1-TI1</td>
<td>O50O2:H18</td>
<td>963</td>
<td>Non AIEC</td>
<td>1.43±0.63</td>
<td>0.03±0.00</td>
<td>1163.70±203.59</td>
</tr>
<tr>
<td>CD 1IM 3</td>
<td>O83:H6</td>
<td>2622</td>
<td>Non-AIEC</td>
<td>3.47±0.27</td>
<td>0.02±0.00</td>
<td>1163.70±203.59</td>
</tr>
<tr>
<td>CD 34 LN</td>
<td>O1:H7</td>
<td>95</td>
<td>Non-AIEC</td>
<td>0.14±0.06</td>
<td>0.02±0.00</td>
<td>458.48±510.69</td>
</tr>
<tr>
<td>CD 62 LN</td>
<td>O2:H4</td>
<td>95</td>
<td>AIEC</td>
<td>4.68±2.07</td>
<td>0.33±0.06</td>
<td>1686.38±307.68</td>
</tr>
<tr>
<td>H001</td>
<td>O8:H10</td>
<td>681</td>
<td>Non-AIEC</td>
<td>1.19±0.02</td>
<td>0.02±0.00</td>
<td>708.12±55.32</td>
</tr>
<tr>
<td>H020</td>
<td>O2:H1</td>
<td>73</td>
<td>Non-AIEC</td>
<td>1.33±0.07</td>
<td>0.08±0.04</td>
<td>1009.20±137.15</td>
</tr>
<tr>
<td>H223</td>
<td>O2:H6</td>
<td>141</td>
<td>AIEC</td>
<td>9.03±2.55</td>
<td>0.46±0.23</td>
<td>435.23±108.95</td>
</tr>
<tr>
<td>H252</td>
<td>O1:H7</td>
<td>95</td>
<td>Non-AIEC</td>
<td>3.47±0.20</td>
<td>0.09±0.04</td>
<td>74.23±30.72</td>
</tr>
<tr>
<td>H263</td>
<td>O2:H7</td>
<td>95</td>
<td>Non-AIEC</td>
<td>0.30±0.03</td>
<td>0.01±0.00</td>
<td>398.51±93.69</td>
</tr>
<tr>
<td>H296</td>
<td>O2:H7</td>
<td>95</td>
<td>Non-AIEC</td>
<td>0.45±0.04</td>
<td>0.00±0.00</td>
<td>792.33±78.74</td>
</tr>
<tr>
<td>H305</td>
<td>O18:H7</td>
<td>95</td>
<td>AIEC</td>
<td>27.52±6.10</td>
<td>0.63±0.24</td>
<td>730.65±68.44</td>
</tr>
<tr>
<td>H397</td>
<td>O1:H7</td>
<td>95</td>
<td>Non-AIEC</td>
<td>0.42±0.09</td>
<td>0.01±0.00</td>
<td>245.22±50.18</td>
</tr>
<tr>
<td>H413</td>
<td>O18:H7</td>
<td>95</td>
<td>Non-AIEC</td>
<td>0.91±0.05</td>
<td>0.02±0.00</td>
<td>717.92±41.86</td>
</tr>
<tr>
<td>H461</td>
<td>O18:H7</td>
<td>95</td>
<td>Non-AIEC</td>
<td>0.40±0.07</td>
<td>0.01±0.00</td>
<td>73.60±19.38</td>
</tr>
<tr>
<td>H504</td>
<td>O18ac:H7</td>
<td>95</td>
<td>AIEC</td>
<td>3.28±1.78</td>
<td>0.28±0.03</td>
<td>330.81±80.68</td>
</tr>
<tr>
<td>H588</td>
<td>H5</td>
<td>126</td>
<td>Non-AIEC</td>
<td>3.38±1.39</td>
<td>0.04±0.00</td>
<td>1145.61±65.45</td>
</tr>
<tr>
<td>H660</td>
<td>O1:H7</td>
<td>95</td>
<td>Non-AIEC</td>
<td>0.52±0.16</td>
<td>0.01±0.00</td>
<td>705.61±145.06</td>
</tr>
<tr>
<td>ED1a</td>
<td>O81:H27</td>
<td>452</td>
<td>Non-AIEC</td>
<td>0.25±0.07</td>
<td>0.05±0.01</td>
<td>68.34±30.95</td>
</tr>
<tr>
<td>LF82</td>
<td>O83:H1</td>
<td>135</td>
<td>AIEC</td>
<td>62.83±5.08</td>
<td>12.23±2.01</td>
<td>227.81±61.37</td>
</tr>
<tr>
<td>K-12</td>
<td>OR:H48</td>
<td>10</td>
<td>Non-AIEC</td>
<td>1.43±0.40</td>
<td>0.02±0.00</td>
<td>22.49±4.34</td>
</tr>
</tbody>
</table>

*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.

¶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.
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, omfF, ugd, dasA, yhfJ, fkpA, btuB, prp and yehH. When the AIEC-ST95 and non-AIEC-ST95 strains were compared, the following genes were identified: yjiM, yjiN, galF, gnd, bchA and ucaK, 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. 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. This cell line actually resulted from hela cell (HeLa) contamination and is of cervical carcinoma origin, not embryonic intestinal origin as originally thought, but is still used as a model for measuring the ability of intestinal bacteria to adhere to and to invade epithelial cells. Other studies of AIEC used cancerous cell lines such as Caco-2 and HT-29 cell lines, derived from colorectal carcinomas, and the HEp-2 cell line, 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 enterohemorrhagic E. coli to colonise the intestinal epithelium. Chassaing et al 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. 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 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 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 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 also observed considerable overlap between the ability of CD and control isolates to replicate within macrophages, and Raisch et al 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. 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.19

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 ENSP 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.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

REFERENCES


---

**Gut microbiota**

---
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

Updated information and services can be found at:
http://gut.bmj.com/content/early/2016/05/19/gutjnl-2015-311059

These include:

References
This article cites 57 articles, 18 of which you can access for free at:
http://gut.bmj.com/content/early/2016/05/19/gutjnl-2015-311059#BIBL

Open Access
This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections
Open access (267)
Crohn's disease (923)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/