

ORIGINAL ARTICLE

Culture independent analysis of ileal mucosa reveals a selective increase in invasive *Escherichia coli* of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum

Martin Baumgart^{1,6}, Belgin Dogan^{1,6}, Mark Rishniw¹, Gil Weitzman², Brian Bosworth², Rhonda Yantiss², Renato H Orsi³, Martin Wiedmann³, Patrick McDonough⁴, Sung Guk Kim⁴, Douglas Berg⁵, Ynte Schukken⁴, Ellen Scherl² and Kenneth W Simpson¹

¹Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA;

²Division of Gastroenterology and Hepatology, Jill Roberts Inflammatory Bowel Disease Center, Weill Cornell Medical College, Cornell University, New York, NY, USA; ³Department of Food Sciences, College of Agriculture and Life Sciences, Cornell University, Ithaca, NY, USA; ⁴Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA and ⁵Department of Molecular Microbiology, School of Medicine, Washington University, St Louis, MO, USA

Intestinal bacteria are implicated increasingly as a pivotal factor in the development of Crohn's disease, but the specific components of the complex polymicrobial enteric environment driving the inflammatory response are unresolved. This study addresses the role of the ileal mucosa-associated microflora in Crohn's disease. A combination of culture-independent analysis of bacterial diversity (16S rDNA library analysis, quantitative PCR and fluorescence *in situ* hybridization) and molecular characterization of cultured bacteria was used to examine the ileal mucosa-associated flora of patients with Crohn's disease involving the ileum (13), Crohn's disease restricted to the colon (CCD) (8) and healthy individuals (7). Analysis of 16S rDNA libraries constructed from ileal mucosa yielded nine clades that segregated according to their origin ($P < 0.0001$). 16S rDNA libraries of ileitis mucosa were enriched in sequences for *Escherichia coli* ($P < 0.001$), but relatively depleted in a subset of Clostridiales ($P < 0.05$). PCR of mucosal DNA was negative for *Mycobacterium avium* subspecies *paratuberculosis*, *Shigella* and *Listeria*. The number of *E. coli in situ* correlated with the severity of ileal disease (ρ 0.621, $P < 0.001$) and invasive *E. coli* was restricted to inflamed mucosa. *E. coli* strains isolated from the ileum were predominantly novel in phylogeny, displayed pathogen-like behavior *in vitro* and harbored chromosomal and episomal elements similar to those described in extraintestinal pathogenic *E. coli* and pathogenic Enterobacteriaceae. These data establish that dysbiosis of the ileal mucosa-associated flora correlates with an ileal Crohn's disease (ICD) phenotype, and raise the possibility that a selective increase in a novel group of invasive *E. coli* is involved in the etiopathogenesis to Crohn's disease involving the ileum.

The ISME Journal (2007) 1, 403–418; doi:10.1038/ismej.2007.52; published online 12 July 2007

Subject Category: microbe-microbe and microbe-host interactions

Keywords: adherent and invasive *E. coli*; dysbiosis; genome subtraction; inflammatory bowel disease; 16S rDNA

Correspondence: Dr KW Simpson, Department of Clinical Sciences, Cornell University, C2001 College of Veterinary Medicine, Ithaca, NY 14850, USA.
E-mail: kws5@cornell.edu

⁶These authors contributed equally to this manuscript.

Received 6 March 2007; revised 19 May 2007; accepted 31 May 2007; published online 12 July 2007

Introduction

Crohn's disease is a chronic debilitating inflammatory bowel disease (IBD) that is prevalent in Europe and North America (Loftus, 2004; Hanauer, 2006). It is widely considered a consequence of uncontrolled intestinal inflammation in response to a combination of elusive environmental, enteric microbial and

immunoregulatory factors in genetically susceptible individuals (Hanauer, 2006; Sartor, 2006). Genetic susceptibility is linked increasingly to defects in innate immunity exemplified by mutations in the innate immune receptor NOD2/CARD15, that in the presence of the enteric microflora may lead to upregulated mucosal cytokine production, delayed bacterial clearance and increased bacterial translocation, thereby promoting and perpetuating intestinal inflammation (Ahmad *et al.*, 2002; Hanauer, 2006; Sartor, 2006; Wehkamp and Stange, 2006). This possibility is supported by studies that have shown the pivotal role of the enteric microflora in the development of IBD in animals with engineered susceptibility (Elson *et al.*, 2005; Kim *et al.*, 2005). However, the relationship of the enteric flora to mucosal inflammation in patients with Crohn's disease is far from clear, with pathogenic bacteria such as *Mycobacterium avium* subspecies *paratuberculosis*, *Listeria*, *Streptococcus* and *Escherichia coli*, (Liu *et al.*, 1995; Shanahan and O'Mahony, 2005; Sartor, 2006; Barnich and Darfeuille-Michaud, 2007), or imbalances in the ratio of beneficial to harmful bacteria termed 'dysbiosis', implicated in the development of inflammation (Tamboli *et al.*, 2004; Seksik *et al.*, 2005).

Advances in molecular microbiology have led to a new awareness of the diversity and complexity of the enteric flora. Only 30% of the fecal flora is considered cultivable, and there is significant variation in the flora in different gastrointestinal segments and luminal contents versus mucosa in healthy individuals (Hayashi *et al.*, 2002; Eckburg *et al.*, 2005; Lepage *et al.*, 2005). The application of labeled oligonucleotide probes targeting the major subgroups of enteric bacteria to intestinal biopsies from patients with Crohn's disease has facilitated culture-independent examination of the spatial distribution of the mucosal flora that is considered to interact most closely with the innate immune system (Kleessen *et al.*, 2002; Swidsinski *et al.*, 2002; Mylonaki *et al.*, 2005; Swidsinski *et al.*, 2005). However, the results have been inconsistent, with some investigators reporting mucosal invasion, (Kleessen *et al.*, 2002; Mylonaki *et al.*, 2005) increased colonization by Enterobacteriaceae (Mylonaki *et al.*, 2005) or the Gamma subdivision of Proteobacteria, Enterobacteriaceae and *Bacteroides/Prevotella* (Kleessen *et al.*, 2002), while others describe the presence of a non-invasive *Bacteroides* biofilm (Swidsinski *et al.*, 2005). This discordance between studies may be due to differences in patient populations and methodology (for example, sites of biopsy and fluorescence *in situ* hybridization (FISH) technique), but it could also reflect a failure to consider Crohn's disease as a group of closely related but distinct disorders, rather than a single disease (Hanauer, 2006; Sartor, 2006). The distribution of intestinal inflammation in patients with Crohn's disease is variable, with disease restricted to the ileum (approximately 33%) or colon (approx-

mately 20%), or involving both the ileum and the colon (approximately 40%) (Ahmad *et al.*, 2002). Differences in the genetic susceptibility (Ahmad *et al.*, 2002) and adaptive immune responses to the enteric microflora (Arnott *et al.*, 2004; Targan *et al.*, 2005) of patients with Crohn's disease involving the ileum, compared to those with CCD, suggest the composition and spatial distribution of the mucosal flora may also vary according to disease phenotype. This notion is supported by the higher prevalence of adherent and invasive *E. coli* (AIEC) strains cultured from the ileal mucosa of patients with Crohn's ileitis, compared to those with CCD (Darfeuille-Michaud *et al.*, 2004; Barnich and Darfeuille-Michaud, 2007), but it has not been substantiated by contemporary culture-independent methodologies.

We address this issue by combining unbiased culture-independent analysis of bacterial diversity with *in situ* localization and culture-based characterization of the ileal mucosa-associated flora of patients with Crohn's disease involving the ileum, patients with CCD and healthy individuals. We show that the ileal mucosa of patients with Crohn's disease involving the ileum frequently harbors higher numbers of *E. coli* and relatively fewer Clostridiales, than patients with CCD and healthy individuals. We demonstrate that the number of *E. coli* visualized in ileal biopsies is related to the severity of ileitis, and discover that invasive *E. coli* strains isolated from the ileum are predominantly novel in phylogeny and harbor chromosomal and episomal elements that are homologous to those described in uropathogenic and avian pathogenic *E. coli*, and pathogenic Enterobacteriaceae. These data establish that dysbiosis of the ileal mucosa-associated flora correlates with an ileal Crohn's disease (ICD) phenotype, and raise the possibility that a selective increase in a novel group of invasive *E. coli* is involved in the etiopathogenesis Crohn's disease involving the ileum. They suggest that an integrated approach that considers an individual's mucosa-associated flora in addition to disease phenotype and genotype may improve outcome.

Materials and methods

Patients and clinical characteristics

Twenty-eight patients were studied: 13 with Crohn's disease involving the ileum (ICD, five with disease restricted to the ileum and eight with ileocolonic involvement), eight with CCD and a normal ileum (CCD) and seven healthy patients undergoing surveillance colonoscopy (Healthy) (Supplementary Table 1). The terminal ileum was intubated as part of the standard of care in all patients. Upon direct visualization of the ileal mucosa, a severity score was recorded on the basis of the Crohn's Disease Endoscopic Index of Severity and the Simple Endoscopic Score for Crohn's Disease (Mary and

Modigliani, 1989; Daperno *et al.*, 2004). Severity was reported as follows: grade 1 = no inflammation; grade 2 = hyperemia; grade 3 = aphthous ulcerations; grade 4 = large ulcers and grade 5 = very large and deep ulcers. To ensure consistency, all endoscopic procedures were performed by the same endoscopist (E Scherl) and scores were recorded in a prospective manner. Ileal biopsies were taken with standard single-use sterile endoscopic forceps. This study was approved by the Cornell University Committee on Human Subjects (Protocol 05-05008). All patients provided signed informed consent to participate and to provide mucosal biopsies to the Tissue Bank (Protocol 0603-859). Patient groups were similar in age (mean \pm s.d.: ICD 45.3 ± 18 , CCD 52.6 ± 18 and Healthy 54.9 ± 11.78) and duration of disease (11.3 ± 11.4 years ICD and 13.1 ± 15.7 years CCD). Patients with ICD had more severe disease than CCD: 5/13 required surgery, 4/13 fistulae, compared with 1/8 surgery and 1/8 fistula. The proportion of patients receiving glucocorticoids, amino salicylic acid (ASA), immunomodulators, biologicals and antimicrobial drugs was not significantly different in ICD and CCD.

Biopsies of ileal mucosa were histologically evaluated for the presence and extent of inflammation by a gastrointestinal pathologist blinded to the other results (Odze *et al.*, 1993). The severity of neutrophilic inflammation was graded as follows: 0 = no inflammation; 1 = neutrophilic cryptitis, limited to <50% of the specimen; 2 = cryptitis with luminal crypt abscesses present in <50% of the specimen and 3 = diffuse (>50% of the specimen) neutrophilic inflammation in multiple crypts with ulceration. The presence or absence of features of chronic injury, such as crypt architectural distortion, pseudopyloric metaplasia and/or granulomas, was also recorded.

16S rDNA libraries

Individual 16S rDNA gene libraries were established for a subset of 20 ileal biopsies that were selected at random from each group enrolled in the study: Healthy (6), ICD (7) and CCD (7). Flash frozen ileal biopsies collected into sterile DNA-free containers and stored at -70°C in the Weill-Cornell Tissue Bank were received frozen in numerically coded cryo-vials and maintained on ice until processing. Biopsies were aseptically transferred from the cryo-vials into a disposable tissue grinder along with 0.25 ml sterile physiological saline solution. Half of the tissue homogenate was evaluated by bacterial culture (described below), and the remainder was used for 16S library construction. DNA was extracted from mucosal homogenates and a 320-bp fragment of bacterial 16S rDNA was amplified by PCR, column purified and cloned into a TA cloning vector (pGEM-T Easy, Promega, Madison, WI, USA), as described previously (Simpson *et al.*, 2006). To remove potential contaminating DNA, sterile

distilled PCR water was filtered through Micron YM-100 filters (Millipore, Bedford, MA, USA) and exposed to UV irradiation for 10 min.

Candidate clones were screened by restriction digest, and clones containing fragments of the correct size were sequenced at the Cornell University BioResource Center, using M13 primers and an ABI 3700 automated DNA sequencer and ABI PRISM-BigDye™ Terminator Sequencing kits with AmpliTaq DNA Polymerase (Applied Biosystems, Foster City, CA, USA). DNA sequences obtained with both forward and reverse primers were cleaned from contaminating vector sequences in Sequencher (Gene Codes Corp., Ann Arbor, MI, USA), proofread against the sequence trace and aligned using MegAlign (DNASTar Inc., Madison, WI, USA). Sequences were submitted to GENBANK (EF205650-EF206282).

Identification of 16S rDNA sequences associated with different source populations

A distance matrix, generated by multiple alignments of 16S rDNA sequences for each isolate using phylogenetic analysis using parsimony (PAUP), was fed into SourceCluster to determine whether isolates from the same group (ICD, CCD, Healthy) clustered by sequence (Nightingale *et al.*, 2006). In order to find specific branches in the phylogenetic tree, in which isolates from a specific group are more abundant or scarce than expected by chance, we used TreeStats. Specifically, TreeStats assesses whether, for each clade in the phylogenetic tree, the observed number of sequences belonging to each group is significantly different from the expected number under the null hypothesis of no clustering among isolates from the same group. A χ^2 goodness-of-fit test was used to compare the observed numbers to the expected numbers in 10 000 permutations of the labeling for each clade (Nightingale *et al.*, 2006). The highest level of significance assigned by Treestats is $P = 0.0001$, which enables Bonferroni correction ($P = 0.05/\text{number of comparisons}$) for 500 multiple comparisons. To further assess the reproducibility of the analysis and to test the resolving power of the 16S analysis, we built a bootstrapped neighbor joining tree (1000 bootstrap replicates). The identity of partial 16S rDNA sequences was determined by comparison to the sequence databases at the Ribosomal Database project (RDPII: <http://rdp.cme.msu.edu/>) and NCBI (BLAST-n).

Quantification of *E. coli* with real-time PCR

Primers (*uidA*-F_5'-TGTGATATCTACCCGCTTCGC-3', *uidA*-R_5'-CAGGAAGTGTTCGCCCTTCA-3') and a probe (5'-/56-FAM/TCGGCATCCGGTCAGTGGCA/3BHQ-1/-3) to *E. coli uidA* were self-designed in the Primer Express 1.0 program for Macintosh OS9 (Applied Biosystems, Foster City, CA, USA). Reactions were performed using AmpliTaq Gold PCR

Master Mix (Applied Biosystems, Foster City, CA, USA) in a 96-well format with 30 μ l reaction volume per well, and for each standard, non-template control or mucosal DNA sample in triplets. Reaction conditions for primers and probes were as follows: 50°C for 2 min, 94°C for 10 min, followed by 40 cycles at 94°C for 20 s, 55°C for 20 s and 72°C for 30 s. A standard curve was generated from DNA isolated from *E. coli* of known concentration determined by quantitative plating. Fluorescence was measured at each 72°C step. To enable normalization to biopsy size, we quantified human 18S rRNA using the control kit RT-CKFT-18S (Eurogentec, Seraing, Belgium), and 18S rRNA values were converted into cell number, according to the manufacturer's instructions. PCR cycling was performed on an ABI Prism Sequence Detection System 7000 (Applied Biosystems, Foster City, CA, USA). Data were analyzed with the Sequence Detection Software 1.2.1 (Applied Biosystems, Foster City, CA, USA) and the number of *E. coli* was expressed as colony-forming unit (CFU)/million human cells.

PCR for *Mycobacterium avium* subspecies *paratuberculosis* (MAP), *Listeria* and *Shigella*

The presence of *Mycobacterium avium* subspecies *paratuberculosis* (MAP), *Listeria* and *Shigella* in DNA extracted from terminal ileal biopsies was examined by PCR with primers directed against IS900, *hlyA* and *ipaH* (Norton *et al.*, 2001; Kim *et al.*, 2002; Martin *et al.*, 2004).

FISH

Formalin-fixed paraffin-embedded histological sections of terminal ileum biopsies from ICD (12), CCD (7) and healthy individuals (8) were mounted on Probe-On Plus slides (Fisher Scientific, Pittsburgh, PA, USA) and evaluated by FISH with probes to all bacteria (EUB338), a subset of the Enterobacteriaceae (1531, 23SrRNA: *E. coli*, *Shigella*, *Salmonella*, *Klebsiella*) and *E. coli/Shigella* (*E. coli*, 16SrRNA), as previously described (Poulsen *et al.*, 1994; Jansen *et al.*, 2000; Simpson *et al.*, 2006). Initial assessment was performed with a combination of EUB-338 (Cy3-5') and a non-binding control probe (non-EUB338: FAM-5'), with subsequent hybridizations employing EUB338-5'/FAM in combination with 1531-Cy3-5') or *E. coli* Cy3-5' (IDT, Coralville, IA, USA). Slides spotted with suspensions of cultured bacteria were used to control probe specificity (Simpson *et al.*, 2006). Sections were examined with an Olympus BX51 epifluorescence microscope. Images were captured with a DP-70 camera and DP-Controller software with Image files processed using DP-Manager (Olympus America, Center Valley, PA, USA). Total bacterial counts were obtained for 10 randomly chosen fields (60 \times) for each section, averaged and expressed as bacteria/mm².

Culture of ileal mucosa-associated bacteria

Frozen ileal biopsies of ICD (13), CCD (8) and healthy controls (7) were ground as described above, and half of the tissue homogenate was used to inoculate trypticase soy agar with 5% sheep blood, Columbia colistin and nalidixic acid (CNA) agar and Gram-negative broth (GN). All media were incubated at 37°C for 18–24 h in 6% CO₂, at which time they were screened for target bacteria, and GN broth was subcultured onto Levine eosin methylene blue (EMB) agar, MacConkey agar with 4-methylumbelliferyl- β -D-glucuronide (MUG), brilliant green agar with novobiocin and xylose–lactose–tergitol 4 agar, and incubated at 37°C for 18–24 h, before checking for colonies; the XLT-4 agar was incubated for an additional 18–24 h if no typical colonies were observed. The ground tissue was also inoculated onto *Campylobacter* agar with five antimicrobials and 10% sheep blood, which was incubated in a microaerophilic chamber at 42°C for 48 h before screening for *Campylobacter* species. Anaerobic culture was performed by plating onto lecithin–lactose agar in an anaerobic chamber, and incubation for 18–24 h at 37°C, before screening for colonies. Aerobic bacteria were screened by Gram stain, catalase and oxidase reaction, and then were identified using the computer controlled, automatic, Sensititre System (TREK Diagnostic Systems, Cleveland, OH, USA). Conventional biochemical reactions using standard identification strategies were used as needed to supplement the Sensititre identification system. Anaerobic bacteria were identified using the Wadsworth disk method. Isolates were archived at –80°C, and fresh non-passaged bacteria used for subsequent investigations as required.

Molecular characterization of *E. coli*

On the basis of 16S rDNA sequence analysis, 10–15 individual *E. coli* colonies from each biopsy were screened by random amplified polymorphic DNA-PCR (RAPD-PCR), and representative isolates that differed in overall genotype were selected for subsequent analyses. Bacterial isolates were stored at –80°C, and fresh non-passaged bacteria were used for all investigations. *E. coli* isolates were streaked on Luria–Bertani (LB) agar and a single colony was inoculated into LB broth. Cells were grown overnight at 37°C without shaking.

The genetic diversity of *E. coli* isolates was evaluated by RAPD-PCR with informative primers 1254 and 1283 (Simpson *et al.*, 2006). The major *E. coli* phylogenetic groups (A, B1, B2 and D) (2, 3) were determined by triplex PCR (Clermont *et al.*, 2000). *E. coli* isolates were serotyped for OH antigens and screened by PCR for the presence of genes encoding K99, F1845 and CS31A fimbriae; heat-labile toxin (LT); heat-stable toxins, STa and STb; Shiga-like toxin types I and II, SLTI and SLTII; cytotoxic necrotizing factors 1 and 2 (*cnf1* and *cnf2*); and intimin- γ (*eae*), at the *E. coli* Reference Center at

Penn State University (DebRoy and Maddox, 2001). The presence of *ibeA*, *papC*, *afaB-afaC*, *sfaD-sfaE* and *focG* was determined by PCR (Martin *et al.*, 2004; Moulin-Schouleur *et al.*, 2006).

Multilocus sequence typing (MLST) for seven loci (*aspC*, *clpX*, *fadD*, *icdA*, *lysP*, *mdh*, *uidA*) was performed according to the protocol established by Whittam and co-workers (http://www.shigatox.net/stec/mlst-new/mlst_pcr.html). Column purified PCR amplicons were sequenced at the Cornell University BioResource Center, using forward and reverse PCR primers and an ABI 3700 automated DNA sequencer and ABI PRISM BigDye Terminator Sequencing kits with AmpliTaq DNA Polymerase (Applied Biosystems, Foster City, CA, USA). DNA sequences obtained with both forward and reverse primers were proof-read, and then assembled in SeqMan (DNASTar, Madison, WI, USA). Sequences were aligned using the Clustal-W algorithm in MegAlign (DNASTar, Madison WI, USA), and allele, st7 and clonal group were determined using the web-based software (http://www.shigatox.net/stec/mlst-new/mlst_pcr.html).

Cell lines and culture conditions

Caco-2 cells (ATCC HTB-37) were grown in minimum essential medium (Gibco, Rockville, MD, USA) supplemented with 15% fetal bovine serum (FBS), 1 mM sodium pyruvate and 0.1 mM non-essential amino acids solution (NEAA). HEp-2 cells (ATCC CCL-23) (derived from a human larynx carcinoma) and the murine macrophage-like cell line J774-A1 were grown in Rosewell Park Memorial Institute 1640 (Gibco-Invitrogen, Grand Island, NY, USA) supplemented with 10% FBS. Monolayers of all cell lines were kept at 37°C in 5%CO₂:95% air (vol/vol). The FBS concentration was dropped to 5% before infection assays.

Invasion of cultured intestinal epithelial cells

The invasive abilities of *E. coli* isolates were evaluated in cultured epithelial cells by the gentamicin protection assay. Caco-2 cells were grown in 24-well plates for 7 days ($\sim 5 \times 10^6$) and infected with *E. coli* strains at a multiplicity of infection (MOI) of 20 for 3 h. Intracellular bacteria were determined as described previously (Simpson *et al.*, 2006). Each assay was run in duplicate and repeated at least three times. Invasion was expressed as the total number of CFU/ml recovered per well. A non-invasive *E. coli* strain (DH5 α , and *E. coli* strain LF82, a strain isolated from a patient with ICD in France that displays adherent and invasive behavior in cultured cells, were used as negative and positive controls, respectively.

Involvement of the host cell cytoskeleton in the invasion process was examined using the microfilament inhibitor cytochalasin D and the microtubule inhibitor colchicines, as described previously (Simpson *et al.*, 2006), with some modifications.

HEp-2 cells were grown to confluency in 24-well plates for 2 days. Cytochalasin D (0.5 μ g/ml) or 1 μ g/ml colchicine was added to HEp-2 cells 30 min before addition of bacteria. Cells were infected at an MOI of 20 for 3 h. After the infection period, cells were washed two times in phosphate-buffered saline (PBS) and then incubated for another 1 h in medium containing gentamicin (100 μ g/ml) to kill extracellular bacteria. The number of intracellular bacteria was determined by plating serial dilutions, as described above. Invasion in the presence of each inhibitor was normalized to the no-inhibitor control for each strain, and reported as a percentage of the control invasion level (100%).

Persistence and replication within epithelial cells and macrophages

To determine whether *E. coli* could survive or multiply within Caco-2 cells, the standard invasion assay was modified by further incubation of infected monolayers for up to 24 h. After the invasion and incubation with gentamicin (100 μ g/ml), cells were washed once in PBS and fresh medium containing 15 μ g/ml of gentamicin was added to cells. At 1 and 24 h post-gentamicin treatment, the number of intracellular bacteria was determined as described above. Survival was expressed as the percentage of bacteria present within cells at 24 h compared to the number internalized at 1 h (100%).

Intracellular survival and replication in J774 macrophages was determined as described previously (Glasser *et al.*, 2001), with some modifications. J774 monolayers were infected at an MOI of 20 bacteria per macrophage. After a 2-h incubation at 37°C with 5% CO₂, infected macrophages were washed twice with PBS and fresh cell culture medium containing gentamicin (100 μ g/ml) was added to kill extracellular bacteria. After incubation for another 1 h, the medium was removed and fresh medium containing gentamicin (20 μ g/ml) was added for longer post-infection periods. Cells were washed once with PBS and 1 ml of 1% Triton X-100 (Sigma Chemical Company, St Louis, MO, USA) in deionized water was added to each well for 5 min, to lyse the eukaryotic cells. This concentration of Triton X-100 had no effect on bacterial viability for at least 30 min. Samples were removed, diluted and plated onto LB agar plates to determine the number of CFUs recovered from the lysed monolayers. The number of bacteria surviving the gentamicin-killing assay was determined at 1 and 24 h post-gentamicin treatment. Survival was expressed as the mean percentage of the number of bacteria recovered after 1 h post-infection, defined as 100%.

Genome subtraction and plasmid separation and sequencing

E. coli strains 541-1, 541-15, LF82 and K-12 (MG1655) were grown in 30 ml LB broth media

without antibiotic at 37°C for overnight. High-quality genomic DNA, extracted using a commercial kit (Qiagen Genomic-tip 500, Qiagen Inc, Valencia CA, USA), was subjected to suppressive subtractive hybridization (Akopyants *et al.*, 1998) using a commercial kit (Clontech PCR-select Bacterial Genome Subtraction kit, Clontech, Mountain View, CA, USA), with *E. coli* MG1655 as the driver strain, with the following modifications to the protocol. Briefly, 10 µg DNA from each strain was digested with *Rsa*I and column purified with DNA purification columns (QIAQuick Nucleotide Removal kit, Qiagen Inc, Valencia CA, USA) to obtain 2 µg DNA for hybridization. Nested PCR products were cloned into a TA cloning vector (pGEM-T-Easy, Promega, Madison, WI, USA) and 40 clones were sequenced. Sequences were examined with BLASTn and BLASTX (NCBI, Bethesda MD, USA) for homology to known bacterial genes. Sequences were submitted to GENBANK (EI011446-EI011511).

Plasmids were extracted and purified from *E. coli* strains 541–15 and LF82 using commercial purification kits. Plasmids were randomly digested with *Tsp*5091 and fragments were cloned and sequenced as described above. Sequences were submitted to GENBANK (EI100642-EI100659).

To ascertain if the nucleotide sequences with homology to pMT1, ColV, *ratA* and *hcp* contained in subtractive libraries of *E. coli* LF82 and 541-15 were associated with ICD, or an adherent and invasive pathotype in cultured cells, we screened our collection of 22 strains by PCR (Supplementary Table 2).

Statistical analysis

The statistical analysis for differences in the origin of 16S sequences is described above. Differences in the frequency of 16S sequences in a library, the number of ileal *E. coli* (determined by quantitative PCR and FISH) in Healthy, CCD and ICD, the *in vitro* pathotype of different phylogroups (A, B1, B2, D) and the frequency of virulence genes in different pathotypes were evaluated by Kruskal–Wallis test, with the Mann–Whitney test as a post-test. Correlation analyses between FISH and PCR, and FISH and PCR versus disease activity, were performed with the Spearman Rank Correlation test. Differences in the *in vitro* pathotype, and frequency of virulence genes in *E. coli* strains isolated from normal ileum

(healthy + CCD) and ICD were evaluated by Mann–Whitney test. Differences in the duration of clinical signs and treatment of ICD and CCD were evaluated by Mann–Whitney test. A value of $P < 0.05$ was considered statistically significant.

Results

16S rDNA libraries of Crohn's ileitis are selectively enriched in sequences for E. coli and depleted in a subset of Clostridiales

To create an inventory of the dominant ileal mucosa-associated flora, we sequenced 616 clones of 16S rDNA amplified from ileal DNA of seven patients with ICD (217 clones), six with CCD with normal ileum (177 clones) and seven healthy individuals (222 clones), at an average of 31 clones per biopsy. Sequence analysis, using a statistical approach designed to identify sequences associated with different source populations, identified nine clades that were significantly associated with a defined source ($P < 0.0001$) (Figure 1). Differences in the prevalence of sequences in individual clone libraries from different source populations (Healthy, ICD, CCD) were found for three clades (Figure 1). A fully annotated version of this analysis is presented in Supplementary Figure 1.

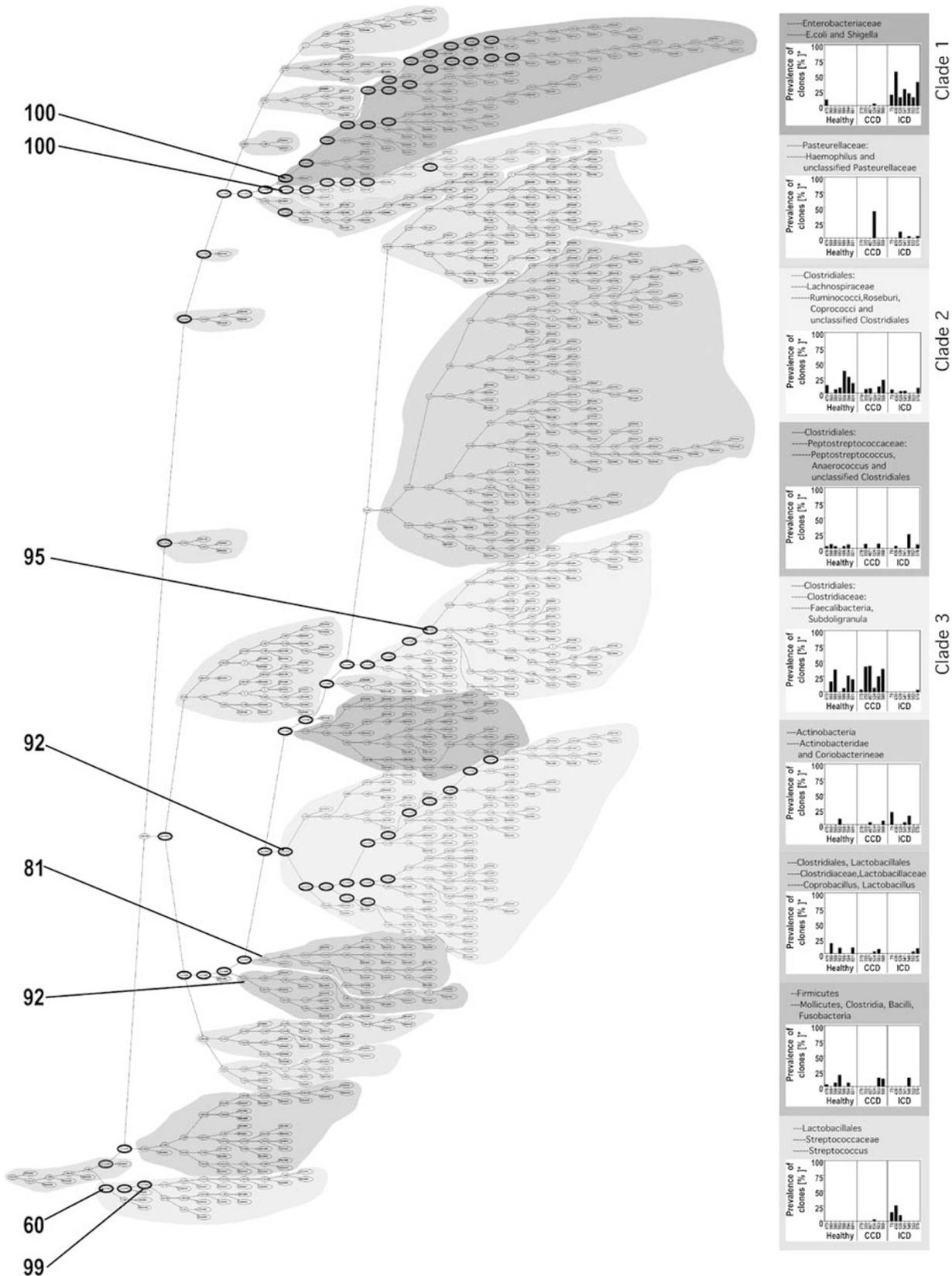
Sequences in the first clade were more prevalent in ICD (26.4%) than ileum from CCD (0.5%, $P = 0.0012$) and Healthy (1.4%, $P = 0.0006$), and were exclusively Enterobacteriaceae, matching *E. coli* and *Shigella* in the NCBI database. PCR to detect *Shigella* (*ipaH*) in ileal DNA was negative, indicating that these sequences originated from *E. coli*. To address the possibility of library bias and to determine whether the increased prevalence of *E. coli* sequences in ICD was relative or absolute, we evaluated the mucosal DNA extracts used for 16S rDNA library construction by quantitative PCR for *E. coli* (*uidA*, normalized to mucosal 18S rDNA). The results (Figure 2) validated our 16S library analysis and provided a clear indication that *E. coli* was significantly more numerous in the ileum of patients with ICD (median 55 352/10⁶ mucosal cells) than CCD (median 10 674/10⁶ mucosal cells, $P = 0.035$) and Healthy (median 410/10⁶ mucosal cells, $P = 0.0175$).

In contrast to the increased prevalence of ICD-associated sequences in clade 1, the second and

Figure 1 Analysis of 16S rDNA sequences from ileal mucosa according to origin. 16S rDNA sequences from Healthy, ICD and CCD were evaluated according to their origin. Nodes marked with a bold oval indicate significant differences according to origin ($P < 0.0001$). The prevalence of sequences in an individual patient clone library for each clade is shown in the bar graphs on the right hand side. Three clades contain sequences that are significantly different in prevalence: Sequences in clade 1 (which showed a bootstrap support of 100% in an NJ tree with 1000 bootstrap replicates) are more prevalent in ICD (26.4%) compared with normal ileum from either CCD (0.5%, $P = 0.0012$) or Healthy (1.4%, $P = 0.0006$), and are exclusively Enterobacteriaceae, matching *E. coli* and *Shigella* in the NCBI database. Sequences in clade 2 (in this clade the branch supported by the last node with a significant *P*-value is supported by bootstrap value of 95%) are less prevalent in libraries of ICD mucosa (3.1% of those in a clone library), compared to Healthy (15.5%, $P = 0.0175$). Sequences in this clade are predominantly Lachnospiraceae, (*Ruminococci*, *Roseburia* and *Coproccoci*). Sequences in clade 3 (bootstrap support of 92%) are less prevalent in ICD (0.4%) compared to Healthy (15%, $P = 0.038$) and CCD (26%, $P = 0.0012$). Sequences in this cluster are predominantly Clostridiales belonging to the genera *Faecalibacteria* and *Subdoligranula*. Bootstrap values from the NJ tree (if > 50) are shown for selected branches. CCD, Crohn's disease restricted to the colon; ICD, ileal Crohn's disease.

third clades contained sequences that were less prevalent in ICD. In clade 2, which contained predominantly Lachnospiraceae, sequences from

ICD mucosa represented only 3.1% of those in a clone library, compared to 15.5% in Healthy ($P=0.0175$) and 7.9% in CCD ($P=0.2949$). In the



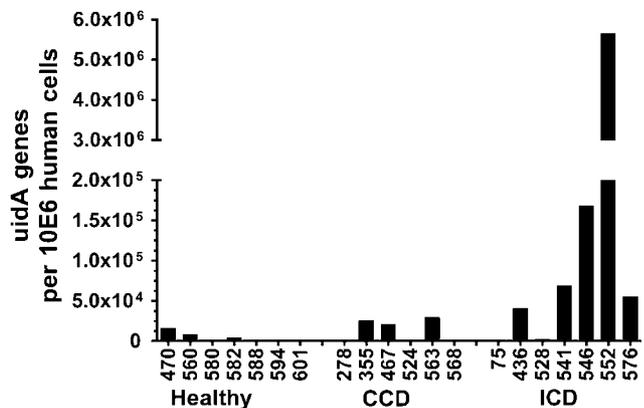


Figure 2 Quantitative PCR for *E. coli uidA* in mucosal DNA from the ileum of healthy individuals, CCD and Crohn's ileitis (ICD). The results of quantitative PCR for *uidA* were normalized to 18S rDNA in mucosal extracts and expressed as *uidA* copies/million cells. CCD, Crohn's disease restricted to the colon; ICD, ileal Crohn's disease.

third clade the prevalence of sequences from Clostridiales belonging to the genus *Faecalibacteria* and *Subdoligranula* were substantially lower in ICD (0.4%), compared to Healthy (15%, $P=0.038$) and CCD (26%, $P=0.0012$).

Analysis of 16S rDNA clone libraries for potential confounding effects of antibiotic consumption showed that the numbers of clones for Enterobacteriaceae (principally sequences resembling *E. coli/Shigella*, clade 1) and Clostridiales (including *Faecalibacteria* clade 3) in libraries constructed from patients with ICD and CCD segregated according to disease status, rather than antibiotic utilization (Supplementary Figure 2).

To address the possibility that pathogenic bacteria linked with Crohn's disease by previous investigators were present in mucosal biopsies at levels that may have precluded their inclusion in 16S rDNA libraries, we PCR amplified mucosal samples with

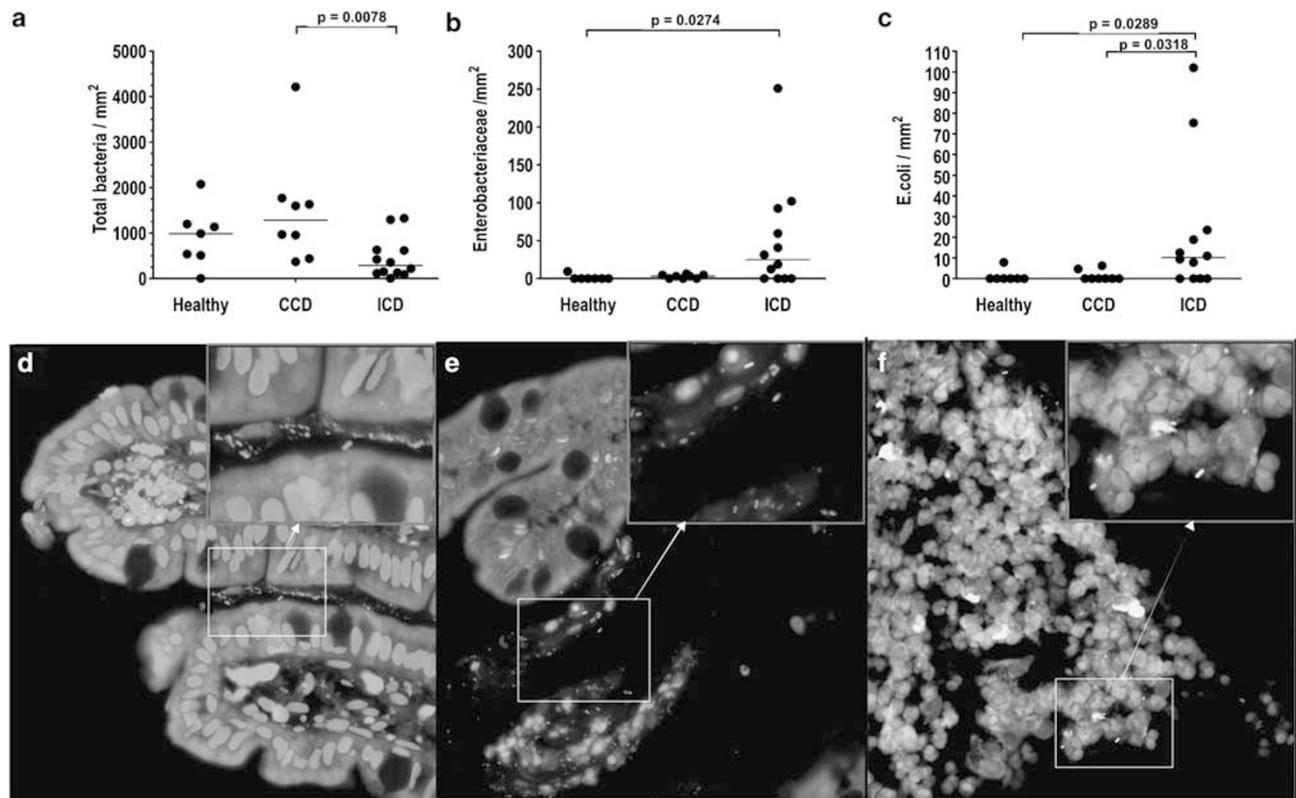


Figure 3 Analysis of the ileal mucosa-associated flora *in situ*. *In situ* hybridization (FISH) with oligonucleotide probes against a subset of Enterobacteriaceae (1531, 23srRNA), *E. coli/Shigella* (16srRNA) and bacteria in general (EUB338) was used to examine the number and spatial distribution of intact bacteria within the ileal mucosa of healthy individuals (Healthy), and patients with CCD and ICD. **(a)** Total bacteria (Eub 338) were more numerous in the ileal mucosa of CCD than ICD ($P=0.0078$). Bar = median. **(b)** Enterobacteriaceae (1531) were more numerous in ICD mucosa (8.7% of mucosal bacteria) than Healthy ileal mucosa (less than 0.2% of mucosal bacteria) were Enterobacteriaceae: $P=0.0274$). Bar = median. **(c)** *E. coli* were more numerous in ICD mucosa (3.5% of mucosal bacteria) than the ileal mucosa of Healthy ($P=0.0289$) and CCD ($P=0.0318$). Bar = median. **(d)** FISH with probe FAM-EUB338 (green) shows that bacteria in normal ileal mucosa are localized predominantly in the superficial mucus covering a villus. DAPI-stained DNA (blue). Bacteria are 2–3 μ long, original magnification, $\times 600$. **(e)** FISH with Cy3-Enterobacteriaceae-1531 (red) and FAM-EUB338 (green) shows a mixed population of Enterobacteriaceae (orange/yellow) and other bacteria (green) on the mucosal surface. DAPI-stained DNA (blue). Bacteria are 2–3 μ long, original magnification, $\times 600$. **(f)** FISH with Cy3-*E. coli/Shigella* (red) and FAM-EUB338 (green) shows *E. coli* (orange/yellow) on and within the eroded ileal mucosa. Intramucosal *E. coli* were observed in 4/12 ICD, but were absent in Healthy and CCD. DAPI-stained DNA (blue). Bacteria are 2–3 μ long, original magnification, $\times 600$. CCD, Crohn's disease restricted to the colon; DAPI, 4,6-diamidino-2-phenylindole; FISH, fluorescence *in situ* hybridization; ICD, ileal Crohn's disease.

primers specific to MAP and *Listeria*. All mucosal samples were negative.

Mucosa-associated Enterobacteriaceae and E. coli are increased in Crohn's ileitis in situ and correlate with disease activity

On the basis of 16S sequence analysis, we used FISH with oligonucleotide probes against a subset of Enterobacteriaceae (1531), *E. coli/Shigella* (*E. coli*) and bacteria in general (EUB338), to examine the spatial distribution of intact mucosa-associated bacteria. Various bacteria (EUB338) were visualized predominantly in the superficial mucus covering a villus, and were less numerous in ICD than CCD (Figure 3, $P=0.0078$). Conversely, ICD mucosa contained significantly more Enterobacteriaceae (8.7% of mucosal bacteria) and *E. coli* (3.5% of mucosal bacteria) than normal ileal mucosa (less than 0.2% of mucosal bacteria were Enterobacteriaceae or *E. coli*) (Figure 3). *E. coli* were most numerous in inflamed and eroded regions, with intramucosal *E. coli* observed in 4/12 ICD but not Healthy or CCD (Figure 3). The number of *E. coli* determined by FISH correlated with the results of quantitative PCR of mucosal DNA for *E. coli uidA* (ρ 0.549, $P=0.0037$). There was no correlation between total bacterial counts and quantitative PCR for *E. coli* (ρ 0.008, $P=0.967$). Additional FISH images are shown in Supplementary Figure 3.

To investigate the relationship of mucosal bacteria to ICD we determined the correlation between the number of bacteria visualized by FISH and the endoscopic and histological scores of ileal disease activity. The number of mucosal Enterobacteriaceae (1531) and *E. coli* visualized by FISH correlated with the severity of ileitis determined by endoscopy (1531 ρ 0.524, $P=0.005$; *E. coli* ρ 0.617, $P=0.0006$) and histology (1531 ρ 0.552, $P=0.0028$; *E. coli* ρ 0.621, $P=0.0006$). In contrast, there was a significant negative correlation between endoscopic disease activity and the total number of mucosa-associated bacteria observed in a section (EUB ρ -0.474, $P=0.0349$).

Ileal E. coli are genetically diverse, belong to novel clonal groups and lack common virulence genes of diarrheagenic E. coli

To determine if the *E. coli* associated with ICD belong to a distinct disease restricted cluster, we screened 205 individual *E. coli* colonies cultured from ileal mucosal homogenates of ICD (110 colonies from 61.5% patients), CCD (70 colonies from 62.5% of patients) and Healthy (25 colonies from 28.5% of patients) for overall genetic diversity and phylogroup (A, B1, B2, D), using RAPD- and triplex-PCR, respectively. Full results of mucosal culture are provided in Figure 4. This approach consolidated the collection of 205 isolates to a group of 22 representative strains: 12 strains from eight patients

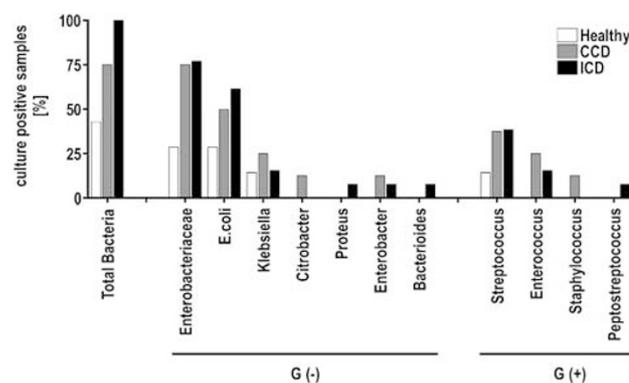


Figure 4 Microbial culture of ileal mucosa. Mucosal homogenates from the ileum of Healthy individuals, CCD and Crohn's ileitis (ICD) were cultured under aerobic and anaerobic conditions. The results are expressed as % of positive cultures in a group. G (-)=Gram negative; G (+)=Gram positive. CCD, Crohn's disease restricted to the colon; ICD, ileal Crohn's disease.

with ICD, eight from five patients with CCD and two from two Healthy individuals.

We then performed additional RAPD-PCR, serotyping, MLST of seven housekeeping genes, and PCR for the presence of genes commonly associated with diarrheagenic *E. coli* (Table 1). The banding patterns generated by RAPD-PCR were unique for 20 of 22 strains, indicating the overall genetic diversity of the population (Figure 5). Two strains from different patients (576-1, 578-1) were identical in RAPD pattern and serotype, and belonged to phylogroup D. Interestingly, MLST indicated that 15 of the 22 strains, regardless of their disease association, belonged to novel clonal groups. We found no evidence for a dominant disease-associated phylogroup or serotype, and of the virulence genes commonly associated with diarrheagenic *E. coli*, only *stx* and *eae* were present in four strains (Table 1). Adhesins associated with extraintestinal pathogenic *E. coli* and APEC (*ibeA* and *papC*) were present in two and four strains respectively (Table 1).

E. coli from healthy and inflamed ileum invade and persist within epithelial cells and replicate in macrophages in vitro

To investigate the potential virulence of *E. coli* strains isolated from ICD mucosa, we studied the ability of our 22 ileal *E. coli* strains to invade intestinal epithelial cells (Caco-2) and replicate within macrophages (J774) *in vitro*. As controls we used a non-pathogenic strain of *E. coli* (DH5 α) and strain LF82, which was isolated from a patient with ICD in France (LF82) and displays 'adherent and invasive' behavior in cultured cells (Boudeau *et al.*, 1999). Four strains with known virulence genes (601-1, 355-1, 584-1, 546-1), and one strain that had intermediary resistance to gentamicin (467-1), were excluded from the culture assays. Of the 17 remaining strains, 10 invaded epithelial cells to a similar degree as LF82 (Figure 6a). Epithelial

Table 1 Characterization of *Escherichia coli* strains isolated from the ileal mucosa of healthy controls, CCD and ICD

Source	Patient ID	Strain	Serotype	Phylogroup	MLST ^a ST7:CG	Virulence genes ^b
Healthy	470	470-1	O1:H12	D	N:0	<i>afa</i> , <i>colV</i>
	601	601-1	O-H30	B1	230:39	<i>Stx1</i> , <i>hcp</i>
CCD	524	524-2	O(140,85):H36	B1	N:0	<i>hcp</i>
		524-6	O(R):H32	A	N:0	<i>colV</i> , <i>hcp</i>
	355	355-1	O86:H18	D	N:0	<i>eae</i> , <i>hcp</i>
		355-5	O43:H-	B1	N:0	<i>hcp</i>
	568	568-2	O7:H+	D	N:0	—
		568-3	O25:H4	B2	34:0	<i>colV</i>
	584	584-1	O132:H2/35	A	171:23	<i>eae</i> , <i>papC</i> , <i>hcp</i>
467	467-1	O21:H+	A	N:0	<i>colV</i> , <i>papC</i> , <i>hcp</i>	
ICD	°532	532-9	O4:H56	B2	28:38	<i>ratA</i> , <i>colV</i>
	°541	541-P6	O8:H23	B1	N:0	<i>colV</i>
		541-P10	O8:H36	D	N:0	<i>ibeA</i> , <i>colV</i>
		541-1	O-H18	B1	N:0	<i>hcp</i>
		541-15	O21:H33	A	N:0	<i>hcp</i>
	°552	552-P4	O116:H9	A	171:23	<i>ratA</i> , <i>hcp</i>
		552-2	O(23 or 73):H28	B1	160:21	<i>hcp</i>
	546	546-1	O-H(6 or 41)	B2	23:7	<i>eae</i> , <i>ibeA</i> , <i>ratA</i> , <i>colV</i>
	576	576-1	O(73 or 77):H18	D	N:0	<i>papC</i> , <i>ratA</i>
	603	603-1	O39:H11	A	N:0	—
	578	578-1	O(73 or 77):H18	D	N:0	<i>papC</i> , <i>ratA</i>
	75	75	O39:H10	A	171:23	<i>hcp</i>

^aAbbreviations: CCD, Crohn's disease restricted to the colon; CG, clonal group; CG0, undefined clonal group; FISH, fluorescence *in situ* hybridization; ICD, ileal Crohn's disease; MLST, multilocus sequence type; N, novel sequence type; ST7, sequence type based on seven housekeeping genes *aspC*, *clpX*, *fadD*, *icdA*, *lysP*, *mdh* and *uidA*.

^bPCR was performed for the following virulence factors: *E. coli* heat-labile toxin (LT), heat-stable toxins (STa and STb), Shiga-like toxins (SLTI and SLTII), cytotoxic necrotizing factors (CNF-1 and CNF-2), the γ variant of the intimin (*eae*) encoding gene, *ibeA*, *papC*, *afaB-afaC*, *sfaD-sfaE*, *focG*, *ratA*, *ColV*, *hcp*, *pMT1*.

^cIntramucosal *E. coli* visualized by FISH.

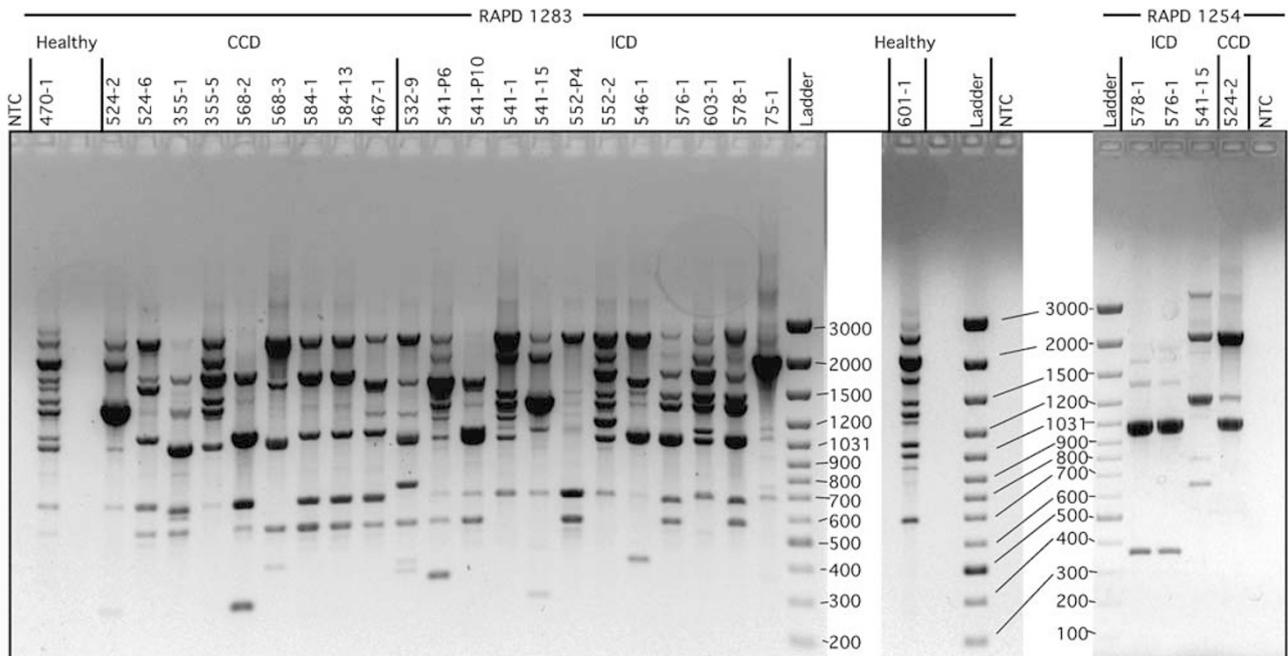


Figure 5 Evaluation of genetic diversity by RAPD-PCR. Strains with different banding patterns are distinct in overall genotype. Two groups of two strains were similar with RAPD primer 1283 (524-2 and 541-15, 576-1 and 578-1). When evaluated with RAPD primer 1254, strains 576-1 and 578-1 were similar in overall genotype (both are D strains and shared serotype O73 or 77:H18), whereas strains 524-2 and 541-15 were different in overall genotype and serotype. RAPD-PCR, random amplified polymorphic DNA-PCR.

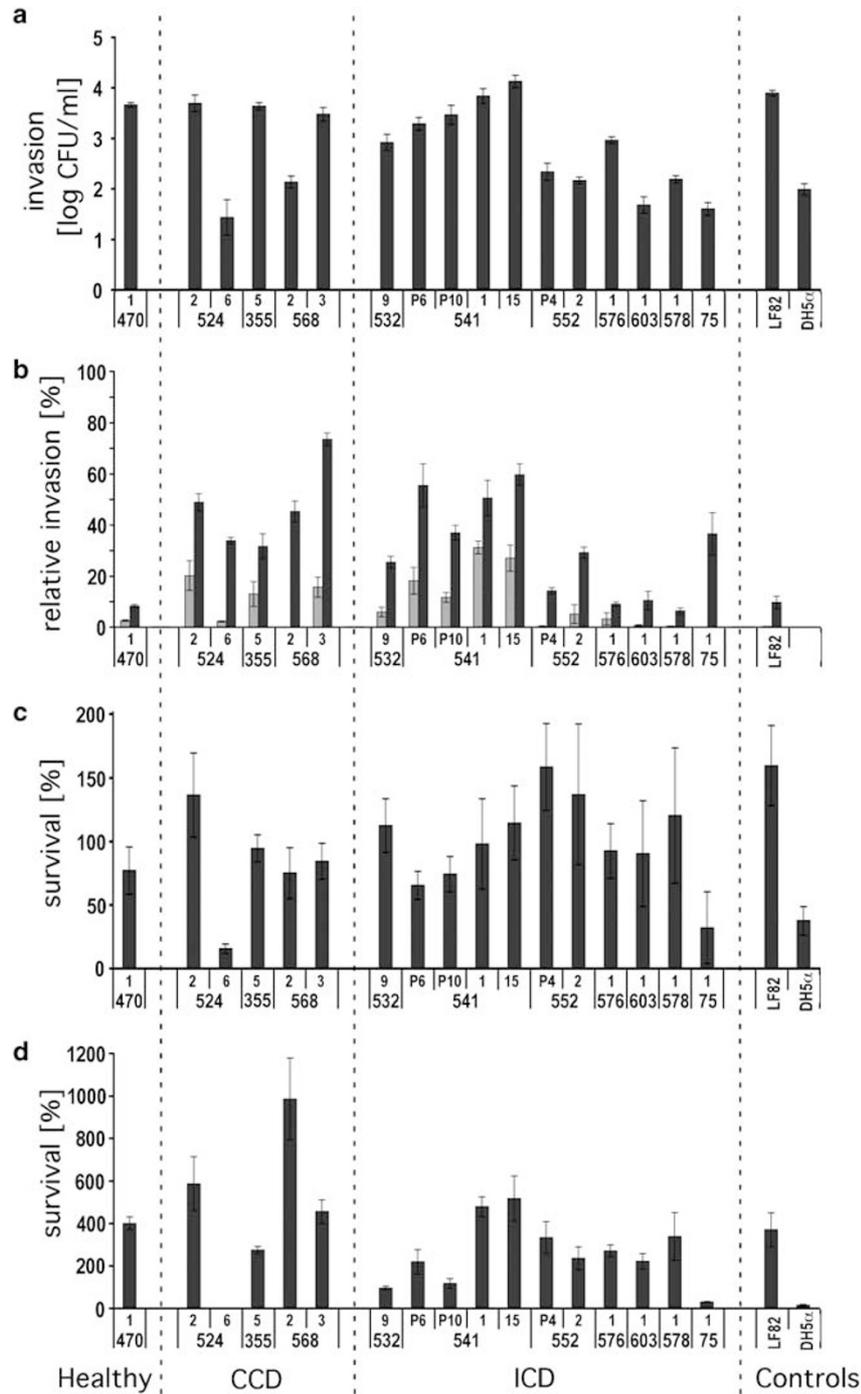


Figure 6 Characterization of ileal *E. coli* in cultured intestinal epithelial cells and macrophages. Bars in the graphs represent the mean \pm s.e. The origin (Healthy, ICD, CCD, control) and individual strain number are shown on the x-axis. Control strain LF82 was isolated from the ileal mucosa of a French patient with ICD, and is the prototype strain for a group known as adherent and invasive *E. coli* (AIEC). (a) Invasion of cultured Caco-2 intestinal epithelial cells by ileal *E. coli* strains and controls. (b) Influence of cytochalasin D (microfilament inhibitor, gray bars) and colchicine (microtubule inhibitor, black bars) on invasion of cultured Caco-2 intestinal epithelial cells. (c) Persistence in cultured Caco-2 intestinal epithelial cells. (d) Persistence and replication in J774 macrophages. Values > 100% are consistent with replication. AIEC, adherent and invasive *E. coli*; CCD, Crohn's disease restricted to the colon; ICD, ileal Crohn's disease; s.e., standard error of the mean.

invasion was reduced between 63% and 99% by the microfilament inhibitor cytochalasin D, and between 22% and 94% by the microtubule inhibitor colchicine, indicating the involvement of host cell cytoskeleton in the invasion process (Figure 6b).

Fifteen strains persisted in epithelial cells more effectively than *E. coli* DH5 α (Figure 6c). Thirteen strains were able to replicate in J774 macrophages (Figure 6d). A similar proportion of strains from normal and ICD mucosa invaded or persisted within

epithelial cells or replicated in macrophages (Figures 6a–d). Strains with an ability to invade epithelial cells and replicate in macrophages were isolated at similar frequencies from patients with ICD (38.5%), CCD (37.5%) and Healthy (14.3%). The ability to invade and persist in cultured cells was independent of phylogroup, as illustrated by strains in phylogroup A that displayed both the lowest and highest levels of invasion.

Genome subtraction reveals homology with extra-intestinal pathogenic E. coli and pathogenic Enterobacteriaceae

To gain additional insight into the phylogeny and virulence of *E. coli* with an adherent and invasive pathotype (AIEC) in cultured cells, we performed genome subtraction on three strains from different phylogroups: two strains isolated from a patient with invasive *E. coli* observed *in situ* (541–1 B1, 541–15 A), and the prototype AIEC strain LF82 (B2). Suppressive subtractive hybridization using the genome sequenced commensal *E. coli* MG1655 as driver yielded 115 genomic fragments with homology to genes that were absent in MG1655. Fifty-one fragments showed homology to genes described in pathogenic *E. coli* (Supplementary Table 3). Thirty-six most closely matched genes in uropathogenic (UPEC) and avian pathogenic (APEC) *E. coli*, and represented at least half of the *E. coli* genes isolated from each strain. The 64 genomic fragments containing non-*E. coli* sequences included 21 with highest homology to pathogenic Enterobacteriaceae such as *Salmonella* (10), *Yersinia* (6), *Shigella* (3) and *Klebsiella* (2) (Supplementary Table 4).

When examined by function, 69 of 115 genomic fragments, representing approximately 50% of the fragments from each strain, encoded putative, hypothetical or novel proteins of as yet unknown function. Of the fragments with homology to known genes, a fragment resembling *ratA* in *Salmonella typhimurium* that is associated with colonization of the cecum and Peyer's patches (Kingsley *et al.*, 2003), and another resembling a hemolysin coregulated protein (*hcp*) associated with virulence in *Vibrio* (Pukatzki *et al.*, 2006), were considered potentially disease relevant. Thirteen sequences for phage or plasmid were detected. Sequences resembling plasmids associated with virulence in *Yersinia pestis* (pMT1) and APEC (ColV) were present in strain LF82, and this was confirmed by the isolation and partial sequencing of a pMT1-like plasmid (Supplementary Table 5). To ascertain if nucleotide sequences with homology to pMT1, ColV, *ratA* and *hcp* were associated with ICD, or an adherent and invasive pathotype in cultured cells, we screened our collection of 22 strains by PCR. pMT1-like sequences were restricted to LF82, whereas *hcp*, ColV and *ratA* were present in 12, eight and five strains, respectively (Table 1). Interestingly, *ratA*

was restricted to adherent and invasive strains isolated from ICD, and was absent in the B1 phylogroup, whereas *hcp* and ColV were evenly distributed in ICD and CCD strains. The presence of these nucleotide sequences in a strain did not correlate with its pathogen-like behavior in cultured cells.

Discussion

The mucosa-associated flora is implicated frequently as a pivotal factor in the development of Crohn's disease, but the specific bacterial characteristics that drive the inflammatory response remain elusive. The lack of success in identifying a specific pathogen or dysbiosis related to Crohn's disease may reflect a failure to consider Crohn's disease as a group of related, but distinct diseases, with similar end points, rather than a single disease (Hanauer, 2006; Sartor, 2006). In the present study, we used a combination of 16S rDNA library analysis, quantitative PCR, FISH and molecular analysis of cultured bacteria, to characterize the ileal mucosa-associated flora of patients with Crohn's disease and healthy individuals, and explored the possibility that the ileal mucosal flora varies according to Crohn's disease phenotype. We have demonstrated that the ileal mucosal flora of patients with ICD is enriched in a novel group of potentially pathogenic *E. coli* and relatively depleted in a subset of Clostridiales, compared to the ileum of healthy individuals and patients with CCD. Our findings establish that the composition of the ileal mucosa-associated flora of patients with Crohn's disease varies according to disease phenotype and provide a rational explanation for the higher prevalence of antibodies directed against *E. coli* outer membrane porin C (OmpC) and flagellin in patients with ICD versus CCD (Arnott *et al.*, 2004; Targan *et al.*, 2005). The relative decreases in a subset of Clostridiales we detected in ICD mirrors recent studies of feces and colonic mucosa from patients with CD, and suggests that this alteration may be a feature of CD in general rather than a specific phenotype (Gophna *et al.*, 2006; Manichanh *et al.*, 2006; Martinez-Medina *et al.*, 2006). These findings in Crohn's disease contrast with the relatively normal numbers of mucosa-associated *Clostridia* reported in patients with ulcerative colitis (Gophna *et al.*, 2006). Studies analyzing feces describe a selective decrease in *Clostridium leptum* group in CD versus a decrease in *Clostridium coccooides* in ulcerative colitis (Sokol *et al.*, 2006).

In the present study, analysis of 16S rDNA libraries with a newly developed statistical approach for assigning sequences according to their origin (Nightingale *et al.*, 2006), and quantitative PCR, were used to guide the selection of oligonucleotide probes for FISH of ileal mucosa. Using an oligonucleotide probe restricted to *E. coli* and

Shigella, and excluding the presence of *Shigella* from mucosal tissues by PCR, we found that *E. coli* represented approximately 3.5% of the EUB338-positive mucosa associated flora and 40% of the Enterobacteriaceae subset recognized by probe 1531 in patients with Crohn's ileitis. Previous studies have chosen FISH probes that target the enteric and fecal flora in general and have inferred the presence of *E. coli* on the basis of hybridization to probe 1531 (Kleessen *et al.*, 2002; Mylonaki *et al.*, 2005; Swidsinski *et al.*, 2005), which is not specific for *E. coli* (Poulsen *et al.*, 1994; Simpson *et al.*, 2006). The relative absence of sequences for *Bacteroides* in 16S libraries of patients with Crohn's ileitis we observed contrasts with the results of previous FISH-based studies reporting *Bacteroides* as the predominant species in the ileal mucosa of patients with Crohn's disease (Kleessen *et al.*, 2002; Swidsinski *et al.*, 2005). Our observation that bacterial invasion was restricted to areas of erosion or ulceration, and the tendency for bacterial numbers to be lower in inflamed than in healthy mucosa, echo previous studies using *in situ* hybridization (Kleessen *et al.*, 2002; Swidsinski *et al.*, 2002). We did not observe the presence of DAPI-positive FISH-negative bacteria in mucosal sections from patients consuming 5-amino salicylic acid (5-ASA) (Swidsinski *et al.*, 2005). Interestingly, this phenomenon appears more frequent in biopsies from patients with ulcerative colitis (UC) (80%) than those with Crohn's disease (CD) (25%) (Swidsinski *et al.*, 2005), hence we may have avoided this effect by focusing on patients with CD, and also sampling the ileum rather than colon where 5-ASA is metabolized to its active form.

Our findings raise the possibility that regional dysbiosis of the mucosa-associated flora is causally related to ileal inflammation in Crohn's disease. The selective increase in *E. coli*, and correlation of *E. coli*, but not bacterial colonization in general, with histological and endoscopic disease activity suggests these bacteria could be involved in the inflammatory process. This notion is supported by the isolation of *E. coli* strains from Crohn's disease mucosa in the present and previous studies that display pathogen-like behavior in cultured cells (Darfeuille-Michaud *et al.*, 2004; Martin *et al.*, 2004). These strains, termed 'adherent and invasive *E. coli*' (AIEC), were originally identified in patients with Crohn's ileitis in France and can persist in cultured macrophages and promote the elaboration of disease relevant cytokines such as tumor necrosis factor α (TNF α) and IL-8 (Boudeau *et al.*, 1999; Glasser *et al.*, 2001; Barnich and Darfeuille-Michaud, 2007). This behavior may account for the presence of *E. coli* antigens and DNA in granulomas of Crohn's disease (Ryan *et al.*, 2004), and the recent association of AIEC with granulomatous colitis in Boxer dogs (Simpson *et al.*, 2006). In the present study, we isolated AIEC strains from 38% of patients with ICD, which is highly consistent with the 36.4%

of patients with ICD described by Darfeuille-Michaud *et al.* (2004).

Bacterial virulence typically is regulated by genes that promote invasion or replication of organisms in target cells (Cossart and Sansonetti, 2004). However, genes of this sort that are absent from commensal *E. coli* have not yet been identified in AIEC strains (Boudeau *et al.*, 2001; Darfeuille-Michaud *et al.*, 2004; Martin *et al.*, 2004; Simpson *et al.*, 2006). The marked heterogeneity in serotype and overall genotype we observed concurs with previous analysis of ribotype (Masseret *et al.*, 2001), and indicates AIEC are not a single virulent clone. Further insight into the relationship of AIEC to *E. coli* as a species was provided by MLST analysis, which revealed that over 75% of the AIEC strains isolated in the present study are phylogenetically distinct from the 679 pathogenic and commensal *E. coli* strains in the EcMLST database. By using PCR-based virulence screening and genome subtraction we discovered that AIEC strains with unique MLST sequences and distinct phylogenetic backgrounds (A, B1, B2) harbor chromosomal and episomal elements that are homologous to those described in UPEC, APEC and pathogenic Enterobacteriaceae such as *Salmonella* and *Yersinia*. The detection of a plasmid resembling pMT1 *Yersinia pestis* in strain LF82, but not other AIEC strains, has important implications, as LF82 has been regarded as the prototype AIEC strain (Boudeau *et al.*, 2001; Barnich and Darfeuille-Michaud, 2007). The possibility that AIEC strains share common pathoadaptive determinants of virulence is supported by the presence in multiple strains of nucleotide sequences with homology to virulence factors in *Salmonella* Typhimurium (*ratA*), APEC (ColV), UPEC (*papC*), meningitis associated *E. coli* (*ibeA*) and *Vibrio* (*hcp*). Nucleotide sequences resembling *ratA* have also been detected in APEC, meningitis associated *E. coli* and UPEC, and may be related to tropism and persistence within the intestinal tract (Schouler *et al.*, 2004). ColV plasmids are important for virulence in APEC and can confer urovirulence on non-pathogenic *E. coli* (Skyberg *et al.*, 2006). *IbeA*, an adhesin originally associated with meningitis associated *E. coli*, is now known to be widely distributed in extraintestinal pathogenic *E. coli* and APEC and is also present in AIEC LF82 (Rodriguez-Siek *et al.*, 2005; Moulin-Schouleur *et al.*, 2006; Simpson *et al.*, 2006). *PapC*, an adhesin associated with pyelonephritis, is also widely distributed in extraintestinal pathogenic *E. coli* and APEC (Rodriguez-Siek *et al.*, 2005), and is present in invasive *E. coli* strains recovered from colonic CD (Martin *et al.*, 2004). Our findings, along with previous results of *fimH* sequencing and virulence gene profiles of LF82 and AIEC from Boxer dogs (Boudeau *et al.*, 2001; Simpson *et al.*, 2006), suggest that AIEC are a novel group of extraintestinal pathogenic *E. coli* associated with chronic intestinal inflammation. While the present study does not

address the role of bacteria in colonic disease, it is noteworthy that adherent and invasive *E. coli* have been associated with Crohn's colitis and colonic cancer (Martin *et al.*, 2004), and *E. coli* of phylogenetic groups B2 and D (these groups contain most of the extraintestinal pathogenic *E. coli*) have recently been linked to colonic inflammation in patients with Crohn's disease and Ulcerative Colitis (Kotlowski *et al.*, 2007). The present study reinforces the high degree of diversity in *E. coli* as a species and its propensity for acquiring DNA from distantly related organisms (Reid *et al.*, 2000; Welch *et al.*, 2002). Genome sequencing will help to resolve whether AIEC strains have acquired common virulence determinants in parallel (Reid *et al.*, 2000), or have adopted the 'mix and match' approach described for UPEC (Brzuszkiewicz *et al.*, 2006), and will aid in the identification of additional candidate genes associated with virulence.

While it is tempting to speculate that AIEC are pathogens rather than harmless commensals, their isolation from healthy and inflamed mucosa indicates that the mere presence of these strains is insufficient to cause disease. The increased numbers of mucosa-associated and invasive *E. coli* relative to other bacteria we observed in patients with Crohn's ileitis suggest that AIEC are opportunistic pathogens that can exploit the mucosal environment of a Crohn's susceptible individual, for example, reduced defensin production and delayed bacterial clearance associated with polymorphisms in NOD2/CARD15 (Sartor, 2006; Wehkamp and Stange, 2006). Alternatively, *E. coli* proliferation may be a consequence of depletion of normal flora such as *Faecalibacterium*, or alterations in bacterial products like butyrate, which is important for colonic health and ileal regeneration (Pryde *et al.*, 2002; Bartholome *et al.*, 2004), or a combination of these possibilities. Resolving these issues may provide unique insights into the biology of Crohn's disease and opportunities for therapeutic intervention.

We conclude that the ileal mucosa of patients with Crohn's disease involving the ileum harbors higher numbers of a phylogenetically novel group of invasive *E. coli* and relatively fewer Clostridiales than the ileal mucosa of patients with CCD, and healthy individuals. Our findings establish that dysbiosis of the ileal mucosal flora correlates with an ICD phenotype, and raise the possibility that a selective increase in a novel group of invasive *E. coli* is involved in the etiopathogenesis of Crohn's ileitis. They suggest that an integrated approach that considers an individual's mucosa-associated flora in addition to disease phenotype and genotype may improve outcome.

Acknowledgements

K Simpson is supported by a grant from the US Public Health Service (DK002938). This study was funded

in part by a seed grant from Cornell University. The authors thank Drs Darfeuille-Michaud and Whittam for *E. coli* strains, Drs Russell and McGregor for helpful discussions, and Francis Davis and Suzanne Klaessig for technical support.

References

- Ahmad T, Armuzzi A, Bunce M, Mulcahy-Hawes K, Marshall SE, Orchard TR *et al.* (2002). The molecular classification of the clinical manifestations of Crohn's disease. *Gastroenterology* **122**: 854–866.
- Akopyants NS, Fradkov A, Diatchenko L, Hill JE, Siebert PD, Lukyanov SA *et al.* (1998). PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. *Proc Natl Acad Sci USA* **95**: 13108–13113.
- Arnott ID, Landers CJ, Nimmo EJ, Drummond HE, Smith BK, Targan SR *et al.* (2004). Sero-reactivity to microbial components in Crohn's disease is associated with disease severity and progression, but not NOD2/CARD15 genotype. *Am J Gastroenterol* **99**: 2376–2384.
- Barnich N, Darfeuille-Michaud A. (2007). Adherent-invasive *Escherichia coli* and Crohn's disease. *Curr Opin Gastroenterol* **23**: 16–20.
- Bartholome AL, Albin DM, Baker DH, Holst JJ, Tappenden KA. (2004). Supplementation of total parenteral nutrition with butyrate acutely increases structural aspects of intestinal adaptation after an 80% jejunioileal resection in neonatal piglets. *JPEN J Parenter Enteral Nutr* **28**: 210–222; discussion 222–223.
- Boudeau J, Barnich N, Darfeuille-Michaud A. (2001). Type 1 pili-mediated adherence of *Escherichia coli* strain LF82 isolated from Crohn's disease is involved in bacterial invasion of intestinal epithelial cells. *Mol Microbiol* **39**: 1272–1284.
- Boudeau J, Glasser AL, Masseret E, Joly B, Darfeuille-Michaud A. (1999). Invasive ability of an *Escherichia coli* strain isolated from the ileal mucosa of a patient with Crohn's disease. *Infect Immun* **67**: 4499–4509.
- Brzuszkiewicz E, Bruggemann H, Liesegang H, Emmerth M, Olschlager T, Nagy G *et al.* (2006). How to become a uropathogen: comparative genomic analysis of extra-intestinal pathogenic *Escherichia coli* strains. *Proc Natl Acad Sci USA* **103**: 12879–12884.
- Clermont O, Bonacorsi S, Bingen E. (2000). Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl Environ Microbiol* **66**: 4555–4558.
- Cossart P, Sansonetti PJ. (2004). Bacterial invasion: the paradigms of enteroinvasive pathogens. *Science* **304**: 242–248.
- Daperno M, D'Haens G, Van Assche G, Baert F, Bulois P, Maunoury V *et al.* (2004). Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD. *Gastrointest Endosc* **60**: 505–512.
- Darfeuille-Michaud A, Boudeau J, Bulois P, Neut C, Glasser AL, Barnich N *et al.* (2004). High prevalence of adherent-invasive *Escherichia coli* associated with ileal mucosa in Crohn's disease. *Gastroenterology* **127**: 412–421.
- DeRoy C, Maddox CW. (2001). Identification of virulence attributes of gastrointestinal *Escherichia coli* isolates of veterinary significance. *Anim Health Res Rev* **2**: 129–140.

- Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M *et al.* (2005). Diversity of the human intestinal microbial flora. *Science* **308**: 1635–1638.
- Elson CO, Cong Y, McCracken VJ, Dimmitt RA, Lorenz RG, Weaver CT. (2005). Experimental models of inflammatory bowel disease reveal innate, adaptive, and regulatory mechanisms of host dialogue with the microbiota. *Immunol Rev* **206**: 260–276.
- Glasser AL, Boudeau J, Barnich N, Perruchot MH, Colombel JF, Darfeuille-Michaud A. (2001). Adherent invasive *Escherichia coli* strains from patients with Crohn's disease survive and replicate within macrophages without inducing host cell death. *Infect Immun* **69**: 5529–5537.
- Gophna U, Sommerfeld K, Gophna S, Doolittle WF, Veldhuyzen van Zanten SJ. (2006). Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. *J Clin Microbiol* **44**: 4136–4141.
- Hanauer SB. (2006). Inflammatory bowel disease: epidemiology, pathogenesis, and therapeutic opportunities. *Inflamm Bowel Dis* **12**(Suppl 1): S3–S9.
- Hayashi H, Sakamoto M, Benno Y. (2002). Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods. *Microbiol Immunol* **46**: 535–548.
- Jansen GJ, Mooibroek M, Idema J, Harmsen HJ, Welling GW, Degener JE. (2000). Rapid identification of bacteria in blood cultures by using fluorescently labeled oligonucleotide probes. *J Clin Microbiol* **38**: 814–817.
- Kim SC, Tonkonogy SL, Albright CA, Tsang J, Balish EJ, Braun J *et al.* (2005). Variable phenotypes of enterocolitis in interleukin 10-deficient mice mono-associated with two different commensal bacteria. *Gastroenterology* **128**: 891–906.
- Kim SG, Shin SJ, Jacobson RH, Miller LJ, Harpending PR, Stehman SM *et al.* (2002). Development and application of quantitative polymerase chain reaction assay based on the ABI 7700 system (TaqMan) for detection and quantification of *Mycobacterium avium* subsp. *paratuberculosis*. *J Vet Diagn Invest* **14**: 126–131.
- Kingsley RA, Humphries AD, Weening EH, De Zoete MR, Winter S, Papaconstantinopoulou A *et al.* (2003). Molecular and phenotypic analysis of the CS54 island of *Salmonella enterica* serotype *typhimurium*: identification of intestinal colonization and persistence determinants. *Infect Immun* **71**: 629–640.
- Kleessen B, Kroesen AJ, Buhr HJ, Blaut M. (2002). Mucosal and invading bacteria in patients with inflammatory bowel disease compared with controls. *Scand J Gastroenterol* **37**: 1034–1041.
- Kotlowski R, Bernstein CN, Sepehri S, Krause DO. (2007). High prevalence of *Escherichia coli* belonging to the B2+D phylogenetic group in inflammatory bowel disease. *Gut* **56**: 669–675.
- Lepage P, Seksik P, Sutren M, de la Cochetiere MF, Jian R, Marteau P *et al.* (2005). Biodiversity of the mucosa-associated microbiota is stable along the distal digestive tract in healthy individuals and patients with IBD. *Inflamm Bowel Dis* **11**: 473–480.
- Liu Y, van Kruiningen HJ, West AB, Cartun RW, Cortot A, Colombel JF. (1995). Immunocytochemical evidence of *Listeria*, *Escherichia coli*, and *Streptococcus* antigens in Crohn's disease. *Gastroenterology* **108**: 1396–1404.
- Loftus Jr EV. (2004). Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterology* **126**: 1504–1517.
- Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L *et al.* (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55**: 205–211.
- Martin HM, Campbell BJ, Hart CA, Mpofo C, Nayar M, Singh R *et al.* (2004). Enhanced *Escherichia coli* adherence and invasion in Crohn's disease and colon cancer. *Gastroenterology* **127**: 80–93.
- Martinez-Medina M, Aldeguer X, Gonzalez-Huix F, Acero D, Garcia-Gil LJ. (2006). Abnormal microbiota composition in the ileocolonic mucosa of Crohn's disease patients as revealed by polymerase chain reaction-denaturing gradient gel electrophoresis. *Inflamm Bowel Dis* **12**: 1136–1145.
- Mary JY, Modigliani R. (1989). Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective multicentre study. Groupe d'Etudes Therapeutiques des Affections Inflammatoires du Tube Digestif (GETAID). *Gut* **30**: 983–989.
- Masseret E, Boudeau J, Colombel JF, Neut C, Desreumaux P, Joly B *et al.* (2001). Genetically related *Escherichia coli* strains associated with Crohn's disease. *Gut* **48**: 320–325.
- Moulin-Schouleur M, Schouler C, Tailliez P, Kao MR, Bree A, Germon P *et al.* (2006). Common virulence factors and genetic relationships between O18:K1:H7 *Escherichia coli* isolates of human and avian origin. *J Clin Microbiol* **44**: 3484–3492.
- Mylonaki M, Rayment NB, Rampton DS, Hudspith BN, Brostoff J. (2005). Molecular characterization of rectal mucosa-associated bacterial flora in inflammatory bowel disease. *Inflamm Bowel Dis* **11**: 481–487.
- Nightingale KK, Lyles K, Ayodele M, Jalan P, Nielsen R, Wiedmann M. (2006). Novel method to identify source-associated phylogenetic clustering shows that *Listeria monocytogenes* includes niche-adapted clonal groups with distinct ecological preferences. *J Clin Microbiol* **44**: 3742–3751.
- Norton DM, Scarlett JM, Horton K, Sue D, Thimothe J, Boor KJ *et al.* (2001). Characterization and pathogenic potential of *Listeria monocytogenes* isolates from the smoked fish industry. *Appl Environ Microbiol* **67**: 646–653.
- Odze R, Antonioli D, Peppercorn M, Goldman H. (1993). Effect of topical 5-aminosalicylic acid (5-ASA) therapy on rectal mucosal biopsy morphology in chronic ulcerative colitis. *Am J Surg Pathol* **17**: 869–875.
- Poulsen LK, Lan F, Kristensen CS, Hobolth P, Molin S, Krogfelt KA. (1994). Spatial distribution of *Escherichia coli* in the mouse large intestine inferred from rRNA *in situ* hybridization. *Infect Immun* **62**: 5191–5194.
- Pryde SE, Duncan SH, Hold GL, Stewart CS, Flint HJ. (2002). The microbiology of butyrate formation in the human colon. *FEMS Microbiol Lett* **217**: 133–139.
- Pukatzki S, Ma AT, Sturtevant D, Krastins B, Sarracino D, Nelson WC *et al.* (2006). Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proc Natl Acad Sci USA* **103**: 1528–1533.
- Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS. (2000). Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* **406**: 64–67.

- Rodriguez-Siek KE, Giddings CW, Doetkott C, Johnson TJ, Fakhr MK, Nolan LK. (2005). Comparison of *Escherichia coli* isolates implicated in human urinary tract infection and avian colibacillosis. *Microbiology* **151**: 2097–2110.
- Ryan P, Kelly RG, Lee G, Collins JK, O'Sullivan GC, O'Connell J *et al.* (2004). Bacterial DNA within granulomas of patients with Crohn's disease – detection by laser capture microdissection and PCR. *Am J Gastroenterol* **99**: 1539–1543.
- Sartor RB. (2006). Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. *Nat Clin Pract Gastroenterol Hepatol* **3**: 390–407.
- Schouler C, Koffmann F, Amory C, Leroy-Setrin S, Moulin-Schouleur M. (2004). Genomic subtraction for the identification of putative new virulence factors of an avian pathogenic *Escherichia coli* strain of O2 serogroup. *Microbiology* **150**: 2973–2984.
- Seksik P, Lepage P, de la Cochetiere MF, Bourreille A, Sutren M, Galmiche JP *et al.* (2005). Search for localized dysbiosis in Crohn's disease ulcerations by temporal temperature gradient gel electrophoresis of 16S rRNA. *J Clin Microbiol* **43**: 4654–4658.
- Shanahan F, O'Mahony J. (2005). The mycobacteria story in Crohn's disease. *Am J Gastroenterol* **100**: 1537–1538.
- Simpson KW, Dogan B, Rishniw M, Goldstein RE, Klaessig S, McDonough PL *et al.* (2006). Adherent and invasive *Escherichia coli* is associated with granulomatous colitis in boxer dogs. *Infect Immun* **74**: 4778–4792.
- Skyberg JA, Johnson TJ, Johnson JR, Clabots C, Logue CM, Nolan LK. (2006). Acquisition of avian pathogenic *Escherichia coli* plasmids by a commensal *E. coli* isolate enhances its abilities to kill chicken embryos, grow in human urine, and colonize the murine kidney. *Infect Immun* **74**: 6287–6292.
- Sokol H, Seksik P, Rigottier-Gois L, Lay C, Lepage P, Podglajen I *et al.* (2006). Specificities of the fecal microbiota in inflammatory bowel disease. *Inflamm Bowel Dis* **12**: 106–111.
- Swidsinski A, Ladhoff A, Pernthaler A, Swidsinski S, Loening-Baucke V, Ortner M *et al.* (2002). Mucosal flora in inflammatory bowel disease. *Gastroenterology* **122**: 44–54.
- Swidsinski A, Weber J, Loening-Baucke V, Hale LP, Lochs H. (2005). Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease. *J Clin Microbiol* **43**: 3380–3389.
- Tamboli CP, Neut C, Desreumaux P, Colombel JF. (2004). Dysbiosis in inflammatory bowel disease. *Gut* **53**: 1–4.
- Targan SR, Landers CJ, Yang H, Lodes MJ, Cong Y, Papadakis KA *et al.* (2005). Antibodies to CBir1 flagellin define a unique response that is associated independently with complicated Crohn's disease. *Gastroenterology* **128**: 2020–2028.
- Wehkamp J, Stange EF. (2006). A new look at Crohn's disease: breakdown of the mucosal antibacterial defense. *Ann N Y Acad Sci* **1072**: 321–331.
- Welch RA, Burland V, Plunkett III G, Redford P, Roesch P, Rasko D *et al.* (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA* **99**: 17020–17024.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)