Crohn's Disease May Be Differentiated Into 2 Distinct Biotypes Based on the Detection of Bacterial Genomic Sequences and Virulence Genes Within Submucosal Tissues

Rodrick J. Chiodini, PhD, Scot E. Dowd, PhD,† Brian Davis, MD,‡ Susan Galandiuk, MD,§ William M. Chamberlin, MD,‖ John Todd Kuenstner, MD,¶ Richard W. McCallum, MD,# and Jun Zhang, PhD**††*

Objective: To determine whether bacterial pathogens can be detected within the diseased submucosal tissues of patients with Crohn's disease by molecular techniques independent of cultural methods.

Design: We designed a quantitative polymerase chain reaction to detect 32 virulence genes and transposons within submucosal tissues of patients with Crohn's disease and controls and compared the microbiome of the submucosa with mucosal bacterial populations.

Results: Within submucosal tissues, the bacterial invasion/adherence genes *eaeA* and *invA* were detected in 43% of patients ($P = 0.01$ and 0.008 vs. mucosa and controls, respectively) and the *Mycobacterium*-specific IS*900* and 251F genes detected in 50% of patients ($P = 0.03$ vs. mucosa and controls). These findings were mutually exclusive: invasion/adhesion genes and *Mycobacterium*-associated transposons were not detected in the same patient. Metagenomic sequencing and quantitative polymerase chain reaction results confirmed effective separation of the submucosal and mucosal microbiome and the existence of a submucosal bacterial population within diseased tissues.

Conclusions: This study is the first to examine the microbial populations of submucosal tissues during intestinal disease and provide evidence of a distinct submucosal microbiome and biotypes within Crohn's disease. These data suggests that Crohn's disease may not be a single disease, but a spectrum that can be divided into distinct biotypes based on the presence of invasion/adherence genes or *Mycobacterium*-associated transposons. If corroborated by larger population studies, these findings could revolutionize the diagnosis, management, and treatment of Crohn's disease by the identification of patient biotypes and the application of targeted chemotherapeutic treatments that go beyond supportive in nature.

Key Words: Crohn's disease, inflammatory bowel disease, infectious agents, invasion and adherence genes, biotypes, mycobacteria, paratuberculosis

(*J Clin Gastroenterol* 2013;00:000–000)

Crohn's disease is a chronic, transmural, granulomatous, idiopathic inflammatory bowel disease of unknown etiology. There is currently no cure and conventional therapy is supportive at best.[1] The natural history of the disease is variable, but the majority of patients tend to have a relapsing and remitting course.[2]

The notion that Crohn's disease is an autoimmune disease has long been dismissed. Current concepts focus on a unique combination of genetic predisposition and environmental triggers that result in immunologic dysregulation releasing a cascade of inflammatory mediators that leads to hyperresponsiveness within the gastrointestinal tract.[3] Bacteria and deficiencies in the innate immune response are now thought to play major roles.[3,4] The predominant etiologic theories include: (1) an abnormally permeable mucosal barrier that results in a continuous translocation of commensal bacteria and/or their antigens into the intestinal wall leading to chronic inflammation, commonly referred to as "leaky gut"[5]; (2) a loss in the balance between "protective" and "harmful" intestinal bacterial populations leading to a disruption of the intestinal microbiome and chronic inflammation, commonly referred to as "dysbiosis"[6]; and (3) the existence of an unidentified persistent bacterial pathogen within intestinal tissues, which continuously drives the inflammatory response.[7,8] However, dysbiosis and leaky gut are common manifestations of gastrointestinal upset[9–12] and cannot explain "skip lesions"[13] or secondary disease sites such as colonic and esophageal often observed in Crohn's disease.[14] Only the persistent pathogen theory directly addresses causality and possible "triggers."

As conventional methods have failed to consistently identify any specific bacterial species in Crohn's disease that can be attributed to disease causation,[8] we sought to identify bacterial virulence genes and other species-specific genomic sequences independent of cultural methodologies. Because Crohn's disease is characterized by transmural

Received for publication June 22, 2012; accepted October 26, 2012.
From the *Division of Gastroenterology; #Division of Gastroenterology; **Department of Internal Medicine; ††Department of Anesthesiology; ‡Department of Surgery, Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center, El Paso; †Molecular Research (Mr. DNA), Shallowater, TX; §Department of Surgery, University of Louisville, Louisville, KY; ‖St. Vincent Physician Network, Sisters of Charity of Leavenworth Health System, Gastroenterology Associates, Billings, MT; and ¶Clinical Laboratories, Charleston Area Medical Center, Charleston, WV.
Supported in part by the Texas Tech University Health Sciences Center, the Lizanell and Colbert Coldwell Foundation, the CAMC Health Education and Research Institute Inc., and the Crohn's Disease Initiative Inc.

Presented in part at American College of Gastroenterology Annual Scientific Meeting, Washington, DC, October 28 to November 2, 2011 and the Society for Surgery of the Alimentary Tract (SSAT), Digestive Disease Week (DDW), San Diego, CA, May 19 to 22, 2012.

The authors declare that they have nothing to disclose.
Reprints: Rodrick J. Chiodini, PhD, Department of Internal Medicine, Division of Infectious Diseases, Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center, 4800 Alberta Avenue, El Paso, TX 79905 (e-mail: rod@baystateservices.com).
Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website, www.jcge.com.

inflammation of the gastrointestinal tract, we hypothesized that bacterial populations within submucosal tissues were more relevant to etiopathogenesis than mucosal biopsies or downstream fecal material used in other studies.[15]

Our findings suggest the existence of at least 2 distinct patient populations or biotypes within the Crohn's disease spectrum and the existence of a submucosal microbiome in both health and disease.

MATERIALS AND METHODS

Patient Populations

Full-thickness resected intestinal tissues from patients with Crohn's disease were obtained from the University of Louisville. Similar tissue samples from patients undergoing intestinal surgery for reasons other than inflammatory bowel disease were obtained from the Texas Tech University Health Sciences Center and the University Medical Center and/or the University of Louisville. A total of 14 tissues from patients with Crohn's disease and 6 normal controls were examined. Clinical parameters of the patient populations are presented in Table 1. Institutional review board approval was obtained from all participating Institutions before the recruitment and enrollment of patients.

Tissue Preparation

Tissues were vortexed in phosphate-buffered saline (PBS) and then washed in fresh PBS for 10 minutes to remove surface contamination. The mucosal layer was then manually excised and approximately 200 mg of the dissected submucosal tissue was vortexed and washed again in PBS for 10 minutes, followed by agitation in 1 mM DL-dithiothreitol (DTT) for 15 minutes to dissolve any residual mucus, and another washing in PBS.

After physical and chemical removal of the mucosal tissues, the submucosa was homogenized in 2 mL of water with a tissue homogenizer. 1 mL of the homogenized tissue was transferred to a bead-beating tube and centrifuged at 10,000 *g* for 2 minutes. DNA was extracted from the sediment as described below. The remaining 1 mL of homogenized tissue was frozen at -80°C for future reference.

Mucosal surface DNA was obtained by dissolving the mucous layer physically excised from resected tissues above in 1 mM DTT for 30 minutes and, after removal of the undigested tissue, centrifugation at 10,000 *g* for 3 minutes. DNA was extracted from the sediment as described below.

DNA Extraction

DNA was extracted from the submucosal tissues based on a modification of the Human Microbiome Project's (HMP) protocol.[16] We found the HMP protocol was ineffective in lysing *Mycobacterium paratuberculosis* (the official name of this organism is *Mycobacterium avium* subspecies *paratuberculosis*, commonly abbreviated as MAP. The authors prefer the original approved name, which more accurately reflects its clinical significance and phenotypic properties)[8] and that the HMP methods failed to yield significant quantities of mycobacterial DNA. As a result, we replaced the garnet beads in the HMP protocol with zirconium, incorporated a 3-minute bead-beating step

TABLE 1. Clinical Features and Characteristics of Patient Population

Patient*	Age	Sex	Ethnicity†	Disease Duration	Disease Location‡	Family History§	Type of Disease‖	Previous Surgery	Drugs (Last 6 mo)
CD01	47	M	W	20 y	Ileum	N	Stricturing	Y	6-mercaptopurine, prednisolone
CD02	49	M	W	24 y	Ileocolonic	CD	Stricturing	Y	5-ASA, B-12, budesonide
CD03	36	M	W	26 y	T. ileum	CA	Stricturing	Y	Remicade, 6-mercaptopurine
CD04	35	M	W	13 y	Ileocolonic	N	Fistular	N	Ciprofloxacin
CD05	32	F	W	5 y	T. ileum	N	Stricturing	N	Remicade, 6-mercaptopurine
CD06	58	F	B	44 y	Colon	N	Pancolitis	Y	Ciprofloxacin, inderol, levothyroxine
CD07	38	M	W	19 y	Small bowel	CD	Stricturing	Y	6-mercaptopurine, prednisolone
CD08	40	M	W	9 y	Colon	N	Dysplasia	N	Pentasa
CD09	36	F	W	14 y	T. ileum	N	Stricturing	Y	Prednisolone
CD10	26	F	W	1.5 y	Colon	N	Stricturing	N	5-ASA, prednisolone
CD11	42	M	B	15 y	Ileocolonic	N	Fistular	N	Prednisolone
CD12	52	F	W	14 y	T. ileum	N	Fistular	Y	Budesonide
CD13	24	M	W	6 y	T. ileum	N	Stricturing	N	Asacol
CD14	49	M	W	2 y	T. ileum	N	Stricturing	N	Humira
nIBD01	59	M	A	6 mo	S. colon	N	Cancer	Y	Invanz, entereg
nIBD02	56	M	H	3 mo	S. colon	N	Cancer	N	Invanz
nIBD03	78	M	H	2 mo	S. colon	N	Volvulus	N	Invanz
nIBD04	55	M	H	2 h	S. colon	N	Trauma	N	Invanz
nIBD05	60	M	H	1 mo	S. colon	N	Cancer	N	Invanz
nIBD06	76	F	W	6 mo	S. colon	N	Diverticulitis	N	Align, ciprofloxacin, methscopolamine

*CD indicates Crohn's disease; nIBD, noninflammatory bowel disease control.
†A indicates Asian; B, African American; H, Hispanic; W, white.
‡Primary disease site and site of examination: S. colon indicates sigmoid colon; T. ileum, terminal ileum.
§Family history of bowel disease: CA indicates colon cancer; CD, Crohn's disease; N, none
‖General disease characteristics at time of surgery.

instead of vortexing, added a manual homogenization step using a tissue homogenizer, and protease K digestion for 20 minutes at 65°C before and after bead-beating.

Target Genes and Polymerase Chain Reaction (PCR) Primers

Target organisms and genes were selected based on bacterial agents that have been implicated as causative agents in Crohn's disease in the past[8] or common bacterial causes of chronic diarrhea (Fig. 1). PCR primers were selected from the literature based on specificity determinations in other studies and the sensitivity/primer-dimer characteristics during optimization as defined below, targeting specific gene sequences (Supplementary Table 1 http://links.lww.com/JCG/A63). For the detection of adherent-invasive *Escherichia coli* (AIEC),[17,18] strains NRG 857C (genebank #CP001855.1) and LF82 (genebank #CU651637.1) were compared and the TPR repeat-containing protein (NRG857_14625) gene (tpR) and glycosyl transferase, family 2 protein (NRG857_10350) gene (*glyT*) were selected as specific for AIEC.

Optimization of Primers and Template DNA

Primers identified in Supplementary Table 1 (http://links.lww.com/JCG/A63) were tested in various molarity combinations to optimize the concentrations and ratios that produced the lowest C_T value, greatest sensitivity, lowest tendency of primer-dimer formation, and consistent T_M values from melting curves. To determine the optimal concentration of template DNA from patient submucosal tissues, gradient dilutions of 50, 100, 150, 300, 500 nM, and 1 μM of template DNA per well spiked with 5 genomes of the corresponding bacterial species were tested with the corresponding primer sets and C_T and T_M values compared.

PCR Assay

Real-time quantitative PCR (qPCR) was performed using the Applied BioSystems Power SYBR-Green Master Mix in an Applied BioSystems Viia-7 Real-Time PCR System using 20 μL total reaction mixture in 384-well plates according to the manufacturer's instructions. All plates were inoculated with an EpMotions 7075 robotic liquid handling system.

Data were analyzed using the Applied BioSystems Viia-7 software. All assays were performed in triplicate and included negative controls without patient template DNA (no template controls, NTC), positive controls with 5 bacterial genomes containing the corresponding target gene sequence, and specific primer inhibition controls containing patient template DNA spiked with 10 bacterial genomes containing the corresponding target gene sequence. Each assay required approximately 40 μg of patient DNA.

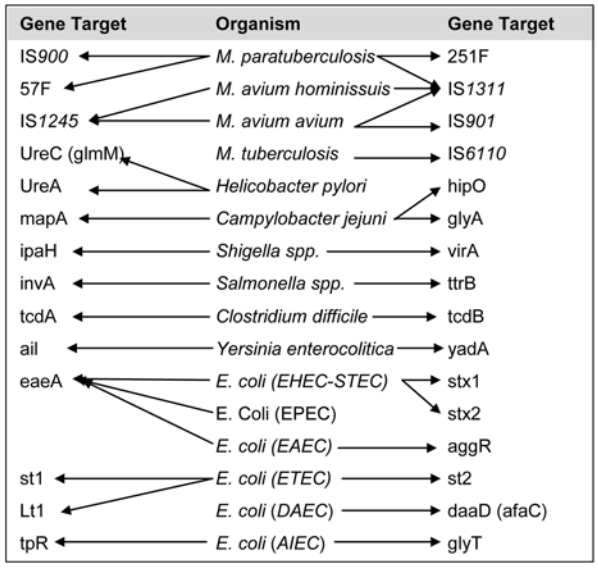

FIGURE 1. Organisms, gene targets, and identification scheme used in quantitative polymerase chain reaction assays.

Verification of Positive qPCR Reactions

All positive qPCR reactions were repeated. Agarose gel electrophoresis was performed on all positive qPCR reactions in 2% agarose. Wells containing DNA bands > 75 bp were excised, purified with a gel extraction kit (GenElute; Sigma) and amplicons sequenced in an Applied Biosystems 3130xl Genetic Analyzer according to the manufacturer's instructions and compared with known sequences within the NCBI database. If any of the NTC controls were positive or if all positive controls, including primer-specific inhibition controls, were not all positive, the corresponding primer set was designated a failure and repeated.

qPCR Design and Analysis

The following criteria were used to designate a positive qPCR reaction: (1) the patient template DNA was positive; (2) all negative controls (NTC) were negative; (3) all positive and template spiked controls were positive; (4) the T_M value of the positive patient DNA was ± 0.5°C of both positive controls; (5) the sample was positive on subsequent qPCR reaction; (6) the amplicon was within 50 bp of the positive controls as determined by gel electrophoresis; and (7) the amplicon sequence was at least 90% homologous with the reference sequence of the gene. qPCR reactions that did not meet these criteria were considered negative.

Quantitation of Target Genes

We used absolute $_{\Delta}C_T$ calculations to quantify DNA as follows: $_{\Delta}C_T$/100 ng × total DNA extracted in ng/tissue weight in grams = genes/g tissue. Total bacterial counts and quantitation of total bacterial communities within submucosal tissues were determined using universal primers as previously described[19] using *E. coli* DNA as $_{\Delta}C_T$ reference and dividing the total genes/g tissue by 7 (copy number of the 16s rDNA in *E. coli*) for bacterial genomes/g tissue. Percent bacterial DNA was calculated using the $_{\Delta}C_T$ of the universal primers times 5.6 fg (average genome size of *E. coli*). Percent bacteria of the total bacterial population containing a specific genomic sequence was calculated as above using the $_{\Delta}C_T$ value of the specific primer set and assuming a single copy of the gene sequence per genome.

Metagenomic Sequencing

rDNA sequencing was performed as previously described by us.[20] The Q25 sequence data derived from the sequencing process were processed using a proprietary analysis pipeline (http://www.mrdnalab.com). Sequences were depleted of barcodes and primers. Short sequences < 150 bp, sequences with ambiguous base calls, and sequences with homopolymer runs exceeding 6 bp were removed. Sequences were then denoized and chimeras

removed. Operational taxonomic units were defined after removal of singleton sequences, clustering at 3% divergence (97% similarity).[20–22] Operational taxonomic units were taxonomically classified using BLAST tool against a curated GreenGenes database[23] and compiled into each taxonomic level into both "counts" and "percentage" files. Count files being the number of sequences and percent being the relative percentage of sequences within each sample.

Statistical Analyses

Data were analyzed using statistical programs inherent within the ViiA-7 analytical software, proprietary analysis software (http://www.mrdnalab.com), Fischer exact test, and/or χ^2 test.

RESULTS

Detection of Specific Sequences in Submucosal Tissues

The following genes were not detected within submucosal tissues of any patient group: IS*1245 of M. avium* complex, *UreA* and *GlyM* (*UreC*) of *Helicobacter pylori*, *mapA*, *hipO*, and *glyA* of *Camplylocater jejuni*, IS*6110* of *Mycobacterium tuberculosis*, *tcdA* and *tcdB* of *Clostridium difficile*, *ipaH* of *Shigella* spp., and *LT1*, *ST2*, *Stx1*, *Stx2*, *aggR*, *afaC* (*daaD*), and *glyT* of the various *E. coli* strains (data not shown).

Virulence genes and/or species-specific bacterial genome sequences were detected in both Crohn's disease and control submucosal tissues (Table 2), suggesting the presence of a submucosal microbiome. The *M. avium* insertion sequence IS*1311* was detected in 43% (6/14) of patients with Crohn's disease and 33% (2/6) controls and the *M. avium* subsp. *avium* IS*901* was detected in 7% (1/14) of Crohn's disease patients and in 17% (1/6) controls. The *M. paratuberculosis* insertion sequence IS*900* was detected in 43% (6/14) patients with Crohn's disease and 17% (1/6) controls, and a unique gene related to lipid metabolism (251F) in *M. paratuberculosis* was concurrently detected in 43% (6/14) of Crohn's disease patients only. The *ttrB* gene of *Salmonella* spp. was also detected in a single patient with Crohn's disease.

Of the *E. coli* pathogenicity genes, the intimin invasion/adherence gene (*eaeA*) was detected in 5 of 14 (36%) Crohn's disease patients but not in any controls, the heat-stable enterotoxin-1 (*st1*) from 1 of 14 (7%) Crohn's disease patients and 1 of 6 (17%) controls, and the *tpr* gene of AIEC from 4 of 14 (29%) patients with Crohn's disease and 2 of 6 controls (33%). In addition to *E. coli*, invasion/adherence genes of *Salmonella* spp. (*invA*) and of *Yersinia* spp. (*yadA*) were detected in 4 of 14 (29%) and in 2 of 14 (14%) Crohn's disease patients respectively, but not in any controls. Multiple adherence/invasion genes were detected in 2 Crohn's disease patients, with both *eaeA* and *invA* detected in 2 patients (14%) and *eaeA* and *yadA* detected in 1 patient (7%).

Except for a single patient with Crohn's disease (resection due to dysplasia) in which no virulence-associated genes were detected, either bacterial *eaeA* or *invA* adherence/invasion genes were detected in 7 of 14 (50%) patients or the IS*900* and 251F genes of *M. paratuberculosis* were detected in 6 of 14 (43%) patients (Table 3). There was a significant difference in the detection of invasion/adherence genes, *eaeA* and *invA*, ($P = 0.01$ and 0.008) and IS*900*/251F genes ($P = 0.03$ and 0.03) within submucosal tissues as compared with mucosa and nIBD controls, respectively. In no case were *M. paratuberculosis*-associated genes (IS*900*/251F) found in the same patient in which adherence/invasion genes were found.

Detection of Specific Sequences in Mucosal Surfaces

The total quantity of DNA obtained from digestion of the intestinal mucosal tissues was insufficient for the examination of all target sequences that required a minimum of 40 μg of DNA. As a result, mucosal DNA was examined only from patients that had positive qPCR results from submucosal tissues.

The detection of bacterial genes within mucosal tissues was generally distinct from that detected within submucosal tissues (Table 3). In most cases, adherence/invasion genes were not detected in mucosal tissues when detected in submucosa of patients with Crohn's disease; in only 1 of 7 submucosal samples from Crohn's disease patients was the *eaeA* adherence/invasion gene detected within the mucosal DNA extracts. In this single case, the amplicon sequences of the *eaeA* genes detected within the mucosa were not identical (97% homology) with that detected within submucosal tissues, suggesting different origins. Genes detected within the submucosa of control tissues were equally found within the mucosa (data not shown).

Discordance of Genomic Detection

In almost all cases, there was discordance in the expected detection of specific genomic sequences as related to a specific bacterial species (Table 2). For example, IS*1311* should have been detected in all tissues in which IS*900* and IS*901* were detected and IS*1245* should have been detected concurrently with the detection of IS*901*. Similarly, the tetrathionate reductase gene (*ttrB*) of *Salmonella* spp. should have been detected in all cases in which the *Salmonella*-specific invasion (*invA*) gene was detected. In addition to discordance in qPCR gene detection, no bacteria that are known to contain IS*900*, 251F, *eaeA*, *invA*, or *yadA* genes were identified by 16S metagenomic sequences (Table 4).

Diversified Submucosal and Mucosal Microbiomes

To further validate the effective separation of mucosal and submucosal bacterial populations, we selected 2 patients with Crohn's disease for metagenomic sequencing of both the submucosal and mucosal microbiome by examination of the bacterial ribosomal DNA (rDNA) genes. One patient was positive for the intimin invasion gene *eaeA* (CD003) and the other for IS*900* and 251F (CD005). Approximately 73% of the rDNA could not be assigned to any specific genus or species; identifiable organisms are shown in Table 4.

Although the mucosa and submucosa were predominately colonized by organisms of the Phyla *Firmicutes* and *Bacteroides*, the relative proportions of these organisms differed between the 2 patients and between mucosa and submucosa in the same patient (Table 4). In patient CD003, which contained the intimin *eaeA* invasion gene, the predominate Phyla in the mucosa were *Firmicutes* (56%) followed by *Bacteroides* (29%) and *Actinobacteria* (15%), whereas in the submucosa, the predominate Phyla were *Bacteroides* (67%) followed by *Actinobacteria* (20%), *Firmicutes* (10%), and *Proteobacteria* (3%). In patient CD005, which contained the *M. paratuberculosis*-associated

TABLE 2. Detection of Virulence Genes and Insertion Sequences in Submucosal Tissues of Patients With Crohn's Disease and Controls*

Patient	IS*900*	251F	*eaeA*	*invA*	IS*1311*	*yadA*	*ttrB*	*ST1*	*tpR*	IS*901*
CD01	+	+	−	−	−	−	−	−	−	−
CD05	+	+	−	−	−	−	−	−	−	−
CD07	+	+	−	−	+	−	−	−	−	−
CD09	+	+	−	−	+	−	−	−	−	−
CD10	+	+	−	−	+	−	−	−	−	−
CD14	+	+	−	−	+	−	−	−	+	−
CD03	−	−	+	−	+	−	−	−	−	−
CD04	−	−	+	−	−	−	−	−	−	+
CD06	−	−	+	−	−	+	+	+	−	−
CD02	−	−	+	+	−	−	−	−	+	−
CD12	−	−	+	+	+		−	−	−	−
CD11	−	−	−	+	−	−	−	−	+	−
CD13	−	−	−	+	−	+	−	−	−	−
CD08	−	−	−	−	−	−	−	−	+	−
nIBD01	−	−	−	−	+	−	−	−	−	−
nIBD02	+	−	−	−	−	−	−	+	−	−
nIBD03	−	−	−	−	+	−	−	−	+	+
nIBD04	−	−	−	−	−	−	−	−	+	−
nIBD05	−	−	−	−	−	−	−	−	−	−
nIBD06	−	−	−	−	−	−	−	−	−	−

*CD indicates Crohn's disease; 251F, sequence of *Mycobacterium paratuberculosis* associated with lipid metabolism; *eaeA*, intimin invasion/adherence gene of *E. coli*; *invA*, inv-spa invasion gene complex of pathogenic *Salmonella* spp.; IS*1311*, insertion sequence of *Mycobacterium avium* complex; IS*900*, insertion sequence of *M. paratuberculosis*; IS*901*, insertion sequence of *M. avium* subsp. *avium*; nIBD, noninflammatory bowel disease; *ST1*, heat-stable enterotoxin-1 of enterotoxigenic *Escherichia coli*; *tpR*, TPR repeat-containing protein of adherent-invasive *E. coli*; *trB*, tetrathionate reductase structural gene of *Salmonella* spp.; *YadA*, adhesion A and invasion gene of pathogenic *Yersinia* spp.

TABLE 3. Detection of Plasmid-mediated Invasion/Adherence Genes and Transposons in Submucosal and Mucosal Tissues From Patients With Crohn's Disease and Controls

	Crohn's Disease					Controls	
	Submucosal Tissue			Mucosal Tissue		Submucosal Tissue	
Gene	Positive	Negative	*P**	Positive	Negative	Positive	Negative
IS*900*	6	8	NS/NS	3	6	1	5
IS*900*/251F	6	8	0.03/0.03	0	6	0	6
eaeA	5	9	NS†/0.05	1†	5	0	6
invA	3	11	NS/NS	0	3	0	6
eaeA or *invA*	8	6	0.01/0.008	1†	8	0	6

*P versus mucosa/control group. One-tailed χ^2 test.

†Sequence of the amplicon did not match the sequences of the genes detected in submucosal tissues of Crohn's disease patient and considered a distinct organism; significant versus mucosa if not considered positive ($P = 0.05$).

NS indicates nonsignificant.

IS*900* and 251F genes, the predominate Phyla in the mucosa were *Firmicutes* (36%) and *Bacteroides* (38%), followed by *Actinobacteria* (12%), *Verrcomicrobia* (7%), and *Proteobacteria* (4%), whereas in the submucosa, the predominate Phyla were *Firmicutes* (65%) followed by *Actinobacteria* (20%), *Bacteroides* (7%), and *Verrucomicrbia* (4%).

In addition to differences in the relative amounts of bacterial Phyla between mucosa and submucosal tissues, different species of bacteria were also found within the submucosa that were not present within or on the mucosa and vice versa (Table 4).

Amplicon Sequencing

The IS*900* and 251F qPCR amplicons were 98% homologous with known sequences of *M. paratuberculosis* in the NCBI database and 99 + % homologous with each other. Minor sequence variations were limited to single-nucleotide variants. The *eaeA* invasion/adherence genes detected were only 96% homologous with known enteropathogenic *E. coli* strains based on the NCBI databases. The *invA* invasion gene, although homologous with each other, only shared 92% identity (similarity) with known *Salmonella* strains identified in NCBI. On the basis of standard metagenomic classification and assignment criteria,[25] the *eaeA* and *invA* genes identified in our studies can only be identified to the unclassified species and unclassified family level, respectively.

Quantitation of Genes Detected and Submucosal Microbiome

The amount of bacterial DNA within submucosal tissues represented only 0.017% of the total DNA extracted. The submucosal tissues from patients with Crohn's disease contained an average of $9.3 \times 10^4 \pm 3.1 \times 10^2$ total bacterial genomes/g of tissue as opposed to an average of $1.2 \times 10^3 \pm 4.3 \times 10^1$ in controls, corroborating increased bacterial translocation. In Crohn's disease patients in which the IS*900*/F251 genes were detected, assuming a single gene copy of the 251F gene per genome, gene counts averaged $6.2 \times 10^3 \pm 4.1 \times 10^2$ or $<2\%$ of the total bacterial population. The *eaeA* invasion/adhesion gene was detected at a frequency of $5.8 \times 10^2 \pm 2.2 \times 10^2$/g of tissue and represented $<0.05\%$ of the total bacterial population count. The *invA* and *yadA* invasion genes were detected at a frequency of $3.4 \times 10^3 \pm 5.1 \times 10^2$ and $4.9 \times 10^3 \pm 1.3 \times 10^2$ genes/g tissue and accounted for $<0.5\%$ of the total bacterial population count, respectively.

Biotypes Correlate With Disease Presentation

The detection of *Mycobacterium*-associated transposons and plasmid-associated adherence-invasion genes were compared with the clinical disease characteristics at time of surgery (Table 1). There was a significant correlation ($P = 0.03$) between the presence of *Mycobacterium*-association transposons with stricturing Crohn's disease (Table 5); IS*900*/251F was only detected within the submucosal tissues of patients with stricturing disease, whereas invasion/adherence genes were equally distributed within all clinical types.

DISCUSSION

This is the first report to examine the bacterial populations and microbiome of submucosal tissues from surgically resected Crohn's disease tissues as opposed to mucosal biopsies or downstream fecal specimens. Our results suggest that Crohn's disease patients may be divided into distinct populations or biotypes: a population containing plasmid-mediated adherence/invasion genes and the other *Mycobacterium*-associated IS*900* and 251F sequences. These biotypes were found to be mutually exclusive: invasion/adherence genes were not found in patients in which IS*900*/251F was detected and vice versa. Although these results need to be corroborated with larger patient populations, the data presented herein also provide the first objective evidence that Crohn's disease may not have a single etiology. The ability to subgroup patient populations may have a profound effect on the diagnosis, management, and ultimate cure of the disease.

The discordance of results from both qPCR and metagenomic sequencing, although unexpected, are not inexplicable. The total amount of bacterial DNA within bulk extracted DNA from submucosal tissues amounted to only approximately 0.017% of the total extracted DNA and sequences detected were 0.05% to 2% of the total bacterial DNA. Thus, the specific sequences detected by qPCR represented only 8.5×10^{-6} to 3.4×10^{-4}% of the total extracted submucosal DNA. These extremely low levels of target DNA sequences likely account for the competitive inhibition observed in our studies and the discordance of results.

The low copy number of target sequences may also account for the inability to detect mycobacterial rDNA.

TABLE 4. Bacterial Populations Detected in the Mucosa and Submucosa of Patients With Crohn's Disease by Metagenomic Sequencing of Bacterial rDNA Genes

	CD003*		CD005†	
Bacterial Population‡	**Mucosa**	**Submucosa**	**Mucosa**	**Submucosa**
Phylum *Firmicutes*				
Staphylococcus epidermidis	19.0	0	1.6	2.0
Veillonella parvula	13.4	0	0	0
Finegoldia magna	8.4	0	0.6	0
Gemella sanguinis	4.2	6.4	0.1	0
Ruminococcus torques	0	0	10.9	22.3
Clostridium bolteae	0	0	5.5	1.0
Blautia product	1.4	2.8	4.5	14.2
Mogibacterium unclassified	0	0	2.8	13.2
Veillonella ratti	0	0	2.2	0.6
Streptococcus oralis	1.4	0	0	0.8
Clostridium difficile	0	0	0.8	0
Anaerococcus hydrogenalis	0	0	0	2.4
Streptococcus parasanguinis	0	1.4	0.9	0.1
Peptoniphilus asaccharolyticus	3.5	0	0	0
Anaerofilum unclassified	0	0	0	0.35
Clostridium innocuum	0	0	1.9	1.9
Blautia unclassified	0	0	1.3	1.9
Clostridium orbiscindens	1.4	0	0.1	0
Streptococcus salivarius	0	0	0.1	
Facklamia sourekii	0	0	0.6	0
Clostridium ramosum	0	0	0	1.2
Roseburia Eubacterium rectal	0	0	0.3	1.8
Anaeroglobus unclassified	0	0	1.3	0
Oceanobacillus picturae	3.5	0	0	0
Percent total *Firmicutes*	56.3	10.6	35.5	64.9
Phylum *Bacteroidetes*				
Parabacteroides johnsonii	12.0	37.0	0	0.4
Bacteroides uniformis	14.8	27.7	0.3	0.1
Bacteroides vulgates	0	0	23.4	2.44
Dechloromonas aromatic	2.1	0	0	1.2
Odoribacter unclassified	0	0	4.3	0.1
Alistipes putredinis	0	0	1.7	1.2
Bacteroides intestinalis	0	0	0.7	0
Ralstonia unclassified	0	2.1	0	0
Parabacteroides distasonis	0	0	3.4	0.2
Dysgonomonas gadei	0	0	0.4	0
Bacteroides ovatus	0	0	2.4	0
Cytophaga unclassified	0	0	0.2	0.1
Bacteroides stercoris	0	0	1.0	1.0
Pedobacter cryoconitis	0	0	0	0.7
Percent total *Bacteroidetes*	28.9	66.7	37.8	7.4
Phylum *Proteobacteria*				
Pelomonas puraquae	0	2.8	0	0
Sutterella parvirubra	0	0	1.7	0.07
Escherichia fergusonii	0	0	2.1	0
Percent total *Proteobacteria*	0	2.8	3.8	0.07
Phylum *Actinobacteria*				
Propionibacterium acnes	9.2	14.2	0	1.2
Collinsella aerofaciens	0	0	6.6	15.6
Corynebacterium amycolatum	2.8	0	0	0
Mycobacterium gordonae	0	0	3.4	0.6
Alloscardovia omnicolens	0	0	0	0.2
Corynebacterium simulans	1.4	2.1	0	0
Eggerthella lenta	0	0	0	0.5
Kocuria rhizophila	0	2.8	0.6	0.3
Bifidobacterium pseudocatenulatum	0	0	1.0	1.3
Propionibacterium granulosum	0	0	0.67	0
Corynebacterium tuberculostearicum	1.4	0.7	0	0
Percent total *Actinobacteria*	14.8	19.9	12.2	19.7
Phylum *Spirochaetes*				
Leptospira licerasiae	0	0	0.2	0.1
Phylum *Fusobacteria*				
Fusobacterium canifelinum	0	0	3.1	2.6
Leptotrichia trevisanii	0	0	0	1.3

TABLE 4. (*continued*)

Bacterial Population‡	CD003*		CD005†	
	Mucosa	Submucosa	Mucosa	Submucosa
Phylum *Verrucomicrobia*				
Akkermansia muciniphila	0	0	7.3	3.8

*CD003: Crohn's disease patient positive for the intimin invasion/adherence gene *eaeA*.
†CD005: Crohn's disease patient positive for the *Mycobacterium*-associated IS*900* and 251F genes.
‡Species designations based on standard metagenomic classification and assignment criteria.[24] Numbers represent relative percent of species in sample.

The bacteria identified by metagenomic sequencing (Table 4) all contain 6 to 7 copies of their 16S rDNA genes. In contrast, *M. paratuberculosis* contains only a single copy of its rDNA gene (believed to be one of the reasons for its slow growth) thereby reducing sensitivity of metagenomic sequencing 6 to 7 times and increased competitive inhibition. Thus, the IS*900*/251F sequences likely represent the detection of *M. paratuberculosis* even in the absence of metagenomic sequence corroboration. Because evidence suggests that IS*900* may not be specific for *M. paratuberculosis*,[8] the detection of IS*900* alone was not considered adequate for identification and the detection of both IS*900* and 251F was required for classification.

The discordance with the invasion/adherence genes *eaeA* and *invA*, however, cannot be explained by competitive inhibition alone. The organisms generally associated with these genomic sequences all contain 6 to 7 rDNA copies and, although competitive inhibition is likely at low target concentrations ($\sim 10^{-7}$% of total DNA), the failure to identify any organism by metagenomic sequencing that are known to contain these plasmids suggests that these virulence genes may exist within yet-to-be-identified species. Discordance of results also supports the suggestion that the *eaeA* and *invA* genes detected may represent yet-to-be-identified bacterial populations within the intestinal microbiome.

The human intestinal tract harbors a dense microbial community ($>10^{12}$ bacteria/g), which has a profound effect on the host's nutrition, physiology, and immune system.[24] Of the well over 1400 different species of bacteria that may reside within the gastrointestinal tract, 80% of species have never been cultured and are unknown.[26] Much of the microbiome's diversity is thought to be attributed to horizontal gene transfer of plasmids.[27] Although gene transfer is generally rare as a result of colonization resistance in the healthy gut that limits direct contact between prospective donor and recipient organisms,[28] altered enterobacterial colonization densities[29] can boost horizontal gene transfer between pathogenic and commensal bacteria.[30]

The specific gene sequences detected by our qPCR methods are either plasmid-mediated or insertion sequences, both of which are known to undergo lateral and recombinant gene transfer.[30,31] It is likely that the *eaeA*, *invA*, and *yadA* invasion/adherence genes detected within submucosal tissues of patients with Crohn's disease have an ancestral origin in *Escherichia*, *Salmonella*, and *Yersinia* spp., respectively, but do not represent recent events. Although the amplicon sequences of the detected genes were >98% homologous to each other, in all cases, amplicon sequences were only between 92% and 96% homologous with the known sequences of their respective species suggesting distant lateral transfer. Using standard metagenomic classification and assignment criteria,[25] the invasion/adherence genes identified in our studies can only be identified to the unclassified species and unclassified family level, suggesting an unrecognized species. Our inability to correlate the detection of invasion/adhesion pathogenicity genes with metagenomic sequencing of the tissues and the discordance of qPCR results based on the NCBI database further suggests the distant gene transfer of *eaeA* and *invA* into commensal bacteria and/or yet-to-be-identified bacterial populations.

The presence of plasmid-mediated invasion/adherence genes could explain high prevalence of phenotypically defined AIEC in certain types of Crohn's disease[17]; our data suggest that the putative specific gene sequences NRG857_14625 and NRG857_10350 are not specific for AIEC. Likewise, the presence of invasion/adherence genes could explain the value of antibiotic treatment in fistular disease as opposed to stricturing disease manifestations.[32,33]

Our results, if corroborated by larger population studies and development of methodologies applicable to a clinical setting, could revolutionize the diagnosis, management, and treatment of Crohn's disease. It would allow the identification of patient subpopulations and biotypes within the Crohn's disease spectrum and the application of targeted chemotherapeutic treatments that go beyond supportive in nature.

TABLE 5. Correlation of Biotype With Clinical Presentation at Time of Surgery

	Stricturing Disease	Fistular Disease
IS*900*/251F*	6	0
eaeA/*invA*	3	4

*$P = 0.03$ for association of IS*900*/251 F with stricturing disease.

REFERENCES

1. Engel MA, Neurath MF. New pathophysiological insights and modern treatment of IBD. *J Gastroenterol.* 2010;45:571–583.
2. Vatn MH. Natural history and complications of IBD. *Curr Gastroenterol Rep.* 2009;11:481–487.
3. Marks DJ, Rahman FZ, Sewell GW, et al. Crohn's disease: an immune deficiency state. *Clin Rev Allergy Immunol.* 2010;38:20–31.
4. Marks DJ. Defective innate immunity in inflammatory bowel disease: a Crohn's disease exclusivity? *Curr Opin Gastroenterol.* 2011;27:328–334.
5. Marteau P. Bacterial flora in inflammatory bowel disease. *Dig Dis.* 2009;27(suppl 1):99–103.
6. Manichanh C, Rigottier-Gois L, Bonnaud E, et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut.* 2006;55:205–211.

7. Behr MA, Schurr E. Mycobacteria in Crohn's disease: a persistent hypothesis. *Inflamm Bowel Dis*. 2006;12:1000–1004.
8. Chiodini RJ, Chamberlin WM, Sarosiek J, et al. Crohn's disease and the mycobacterioses: A quarter century later. Causation or simple association? *Crit Rev Microbiol*. 2012;38:52–93.
9. Cani PD, Delzenne NM. Interplay between obesity and associated metabolic disorders: new insights into the gut microbiota. *Curr Opin Pharmacol*. 2009;9:737–743.
10. De Palma G, Nadal I, Medina M, et al. Intestinal dysbiosis and reduced immunoglobulin-coated bacteria associated with coeliac disease in children. *BMC Microbiol*. 2010;10:63.
11. Cremon C, Carini G, De Giorgio R, et al. Intestinal dysbiosis in irritable bowel syndrome: etiological factor or epiphenomenon? *Expert Rev Mol Diagn*. 2010;10:389–393.
12. Turnbaugh PJ, Ridaura VK, Faith JJ, et al. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med*. 2009;1:6ra14.
13. Makharia GK, Srivastava S, Das P, et al. Clinical, endoscopic, and histological differentiations between Crohn's disease and intestinal tuberculosis. *Am J Gastroenterol*. 2010;105: 642–651.
14. Clarke BW, Cassara JE, Morgan DR. Crohn's disease of the esophagus with esophagobronchial fistula formation: a case report and review of the literature. *Gastrointest Endosc*. 2010;71:207–209.
15. Kang S, Denman SE, Morrison M, et al. Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm Bowel Dis*. 2010;16: 2034–2042.
16. Anonymous. HMP Human Microbiome Project—Core Microbiome Sampling Protocol A. 2010.
17. Nash JH, Villegas A, Kropinski AM, et al. Genome sequence of adherent-invasive *Escherichia coli* and comparative genomic analysis with other *E. coli* pathotypes. *BMC Genomics*. 2010;11:667.
18. Miquel S, Peyretaillade E, Claret L, et al. Complete genome sequence of Crohn's disease-associated adherent-invasive *E. coli* strain LF82. *PLoS One*. 2010;5:e12714.
19. Nadkarni MA, Martin FE, Jacques NA, et al. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology*. 2002;148:257–266.
20. Dowd SE, Callaway TR, Wolcott RD, et al. Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiol*. 2008;8:125.
21. Capone KA, Dowd SE, Stamatas GN, et al. Diversity of the human skin microbiome early in life. *J Invest Dermatol*. 2011;131:2026–2032.
22. Swanson KS, Dowd SE, Suchodolski JS, et al. Phylogenetic and gene-centric metagenomics of the canine intestinal microbiome reveals similarities with humans and mice. *ISME J*. 2011;5:639–649.
23. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol*. 2006;72: 5069–5072.
24. Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol*. 2009;9:313–323.
25. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26:2460–2461.
26. Qin J, Li R, Raes J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464:59–65.
27. Johnson TJ, Thorsness JL, Anderson CP, et al. Horizontal gene transfer of a ColV plasmid has resulted in a dominant avian clonal type of *Salmonella enterica* serovar Kentucky. *PLoS One*. 2010;5:e15524.
28. Scott KP. The role of conjugative transposons in spreading antibiotic resistance between bacteria that inhabit the gastrointestinal tract. *Cell Mol Life Sci*. 2002;59:2071–2082.
29. Pedron T, Sansonetti P. Commensals, bacterial pathogens and intestinal inflammation: an intriguing menage a trois. *Cell Host Microbe*. 2008;3:344–347.
30. Stecher B, Denzler R, Maier L, et al. Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proc Natl Acad Sci USA*. 2012;109:1269–1274.
31. da Silva Rabello MC, de Oliveira RS, Silva RM, et al. Natural occurrence of horizontal transfer of *Mycobacterium avium*-specific insertion sequence IS1245 to *Mycobacterium kansasii*. *J Clin Microbiol*. 2010;48:2257–2259.
32. Hvas CL, Dahlerup JF, Jacobsen BA, et al. Diagnosis and treatment of fistulising Crohn's disease. *Dan Med Bull*. 2011;58:C4338.
33. Felley C, Vader JP, Juillerat P, et al. Appropriate therapy for fistulizing and fibrostenotic Crohn's disease: Results of a multidisciplinary expert panel—EPACT II. *J Crohns Colitis*. 2009;3:250–256.