



(86) Date de dépôt PCT/PCT Filing Date: 2007/05/03
(87) Date publication PCT/PCT Publication Date: 2007/11/08
(85) Entrée phase nationale/National Entry: 2008/10/31
(86) N° demande PCT/PCT Application No.: CA 2007/000740
(87) N° publication PCT/PCT Publication No.: 2007/124591
(30) Priorité/Priority: 2006/05/03 (US60/796,879)

(51) Cl.Int./Int.Cl. *C12Q 1/68* (2006.01),
C12N 1/20 (2006.01), *C12Q 1/04* (2006.01),
G01N 33/53 (2006.01), *G01N 33/566* (2006.01)

(71) Demandeur/Applicant:
UNIVERSITY OF MANITOBA, CA

(72) Inventeurs/Inventors:
KRAUSE, DENIS O., CA;
BERNSTEIN, CHARLES N., CA

(74) Agent: SMART & BIGGAR

(54) Titre : MARQUEURS MICROBIENS DE COLOPATHIE FONCTIONNELLE
(54) Title: MICROBIAL MARKERS OF INFLAMMATORY BOWEL DISEASE

(57) **Abrégé/Abstract:**

The present invention relates to bacterial serine protease autotransporter (SPATE) and Antigen 43 (Ag43) and their various uses relating to inflammatory bowel disease (IBD) , specifically in the diagnosis of IBD and the screening of potential agents for treating IBD . The invention also relates to methods for cultivating and identifying enteric microbes.



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
8 November 2007 (08.11.2007)

PCT

(10) International Publication Number
WO 2007/124591 A1

(51) International Patent Classification:

C12Q 1/68 (2006.01) *G01N 33/53* (2006.01)
C12N 1/20 (2006.01) *G01N 33/566* (2006.01)
C12Q 1/04 (2006.01)

(21) International Application Number:

PCT/CA2007/000740

(22) International Filing Date: 3 May 2007 (03.05.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/796,879 3 May 2006 (03.05.2006) US

(71) Applicant (*for all designated States except US*): **UNIVERSITY OF MANITOBA** [CA/CA]; 631 Drake Centre, Winnipeg, Manitoba R3T 5V4 (CA).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **KRAUSE, Denis, O.** [CA/CA]; c/o UNIVERSITY OF MANITOBA, 631 Drake Centre, Winnipeg, Manitoba R3T 5V4 (CA). **BERNSTEIN, Charles, N.ah** [CA/CA]; c/o UNIVERSITY OF MANITOBA, 631 Drake Centre, Winnipeg, Manitoba R3T 5V4 (CA).

(74) Agent: **SMART & BIGGAR**; P.O. Box 2999, Station D, 900-55 Metcalfe Street, Ottawa, Ontario K1P 5Y6 (CA).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- with amended claims
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: MICROBIAL MARKERS OF INFLAMMATORY BOWEL DISEASE

(57) Abstract: The present invention relates to bacterial serine protease autotransporter (SPATE) and Antigen 43 (Ag43) and their various uses relating to inflammatory bowel disease (IBD), specifically in the diagnosis of IBD and the screening of potential agents for treating IBD. The invention also relates to methods for cultivating and identifying enteric microbes.

WO 2007/124591 A1

MICROBIAL MARKERS OF INFLAMMATORY BOWEL DISEASE

REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 60/796,879, filed May 3, 2006, the content of which is herein incorporated by reference.

FIELD OF INVENTION

The present invention relates to microbial genes of the autotransporter family, specifically the serine protease autotransporter (SPATE) family and the Antigen 43 (Ag 43) family, and their role in inflammatory bowel disease (IBD). The invention also relates to methods for cultivating and identifying microbes of the gut.

BACKGROUND OF THE INVENTION

Inflammatory bowel disease (IBD) is a collective term for ulcerative colitis (UC) and Crohn's disease (CD). These diseases are chronic inflammatory diseases of the digestive tract, potentially leading to severe inflammation, ulceration, and obstruction, the end-point of which may be surgical resection.

IBD is thought to be a result of recognition of a microbial antigen(s) by a dysfunctional immune system in a genetically predisposed host. Many bacteria have been linked to IBD, but specific bacteria may have been missed because not more than 30% of the microbial diversity in the gut can be cultured (refs 8 to 12).

Several bacteria have been implicated in the aetiology of IBD, the most prominent among these being *Mycobacterium paratuberculosis*. Other bacteria that have been associated are members of the Enterobacteriaceae, *Helicobacter pylori*, and *Bacteroides* species.

SUMMARY OF THE INVENTION

We describe here an invention based in part, but is not limited to, the use of culture-independent surveys of microbial diversity as a prelude to targeted cultivation of bacteria. As an example, we describe the use of ribosomal intergenic spacer analysis (RISA) of biopsy tissue to identify nucleic acids that are consistently associated with IBD. From our survey, we found that genes of the microbial serine protease autotransporter (SPATE) family and the Ag43 family are useful markers for diagnosing and prognosing IBD.

We describe a method for diagnosing inflammatory bowel disease (IBD) or determining susceptibility to developing IBD in a subject. The method comprises the step of assaying for SPATE or Ag43 or both in an enteric bacteria-containing sample from the subject, wherein the presence of SPATE or Ag43 or both in the sample indicates that the subject has IBD or is susceptible to developing IBD.

The subject undergoing diagnosis or prognosis for IBD includes those having symptoms of IBD, or are suspected of having IBD. IBD includes ulcerative colitis (UC) and Crohn's disease (CD).

In the diagnostic or prognostic method, the SPATE or Ag43 being assayed may be SPATE nucleic acid or SPATE

polypeptide or Ag43 nucleic acid or Ag43 polypeptide. With respect to detecting SPATE or Ag43 nucleic acid, the method may comprise detecting a region of SPATE or a region of Ag43 which is conserved among enteric bacteria, or detecting a region of
5 SPATE or a region of Ag43 which is conserved among enteric E. coli, or detecting a region of SPATE or a region of Ag43 which is specific to enteric E. coli, or detecting a region of SPATE or a region of Ag43 which is conserved among virulent enteric strains of E. coli, or detecting a region of SPATE or a region
10 of Ag43 which is specific to virulent enteric strains of E. coli, or detecting a region of SPATE or a region of Ag43 which is specific to E. coli of the B2 or D or the B2+D genotype. Polymerase chain reaction (PCR) may be used to detect SPATE or Ag43 nucleic acid. Primer sequences may be designed for PCR
15 amplification of the target sequences in these methods.

In the diagnostic or prognostic method, the SPATE or Ag43 being assayed may be SPATE or Ag43 polypeptide. The method may employ an immunoassay for SPATE or Ag43 polypeptide. For example, the immunoassay may employ an antibody which is
20 immunospecific against SPATE or Ag43 of enteric bacteria, or is immunospecific against SPATE or Ag43 of enteric E. coli, or is immunospecific against SPATE or Ag43 of virulent enteric strains of E. coli, or is immunospecific against SPATE or Ag43 of E. coli of the B2 or D or B2+D genotype. The diagnostic or
25 prognostic method may involve determining the level of SPATE or Ag43 in the sample, in which case a higher level of SPATE or Ag43 compared to a cut-off value would indicate that the subject has IBD or is susceptible to developing IBD.

In the various methods disclosed here which include diagnostic or prognostic methods, the subject includes human or an animal suitable for use in an IBD disease model, e.g. mouse.

In the various methods disclosed here which include
5 diagnostic or prognostic methods, the sample is any sample containing enteric bacteria. Sources for samples include gut tissue biopsy, intestinal mucosa, stool or fecal matter, or intestinal wash. In particular, the sample may be a
colonoscopy tissue biopsy from the lower gastrointestinal (GI)
10 tract.

We also describe a method for evaluating effectiveness of a treatment for inflammatory bowel disease (IBD) in a subject, the method comprising the steps of: (i) determining serine protease autotransporter (SPATE) level or
15 Ag43 level, or both, in an enteric bacteria-containing sample from a treated subject having IBD, and (ii) comparing the SPATE level or the Ag43 level or both from step (i) to SPATE and Ag43 levels determined in an enteric bacteria-containing sample from an untreated subject having IBD. The untreated and treated
20 subjects are either different subjects or are the same subjects before and after undergoing the treatment. A lower level of SPATE or Ag43 or both in treated subjects indicates that the treatment is effective for IBD.

The subject undergoing evaluation for IBD treatment
25 includes those having ulcerative colitis (UC) or Crohn's disease (CD).

In the methods for evaluating IBD treatment, the SPATE or Ag43 level may be determined by assaying for SPATE

nucleic acid or SPATE polypeptide or Ag43 nucleic acid or Ag43 polypeptide. With respect to detecting nucleic acid, the method may comprise detecting a region of SPATE or Ag43 which is conserved among enteric bacteria, or detecting a region of
5 SPATE or Ag43 which is conserved among enteric E. coli, or detecting a region of SPATE or Ag43 which is specific to enteric E. coli, or detecting a region of SPATE or Ag43 which is conserved among virulent enteric strains of E. coli, or detecting a region of SPATE or Ag43 which is specific to
10 virulent enteric strains of E. coli, or detecting a region of SPATE or Ag43 which is specific to E. coli of the B2 or D or the B2+D genotype. Polymerase chain reaction (PCR) may be used to detect SPATE or Ag43 nucleic acid. Primer sequences may be designed for PCR amplification of the target sequences in these
15 methods.

In the methods for evaluating IBD treatment, the SPATE or Ag43 being assayed may be SPATE or Ag43 polypeptide. The method may employ an immunoassay for SPATE or Ag43 polypeptide. For example, the immunoassay may employ an
20 antibody which is immunospecific against SPATE or Ag43 of enteric bacteria, or is immunospecific against SPATE or Ag43 of enteric E. coli, or is immunospecific against SPATE or Ag43 of virulent enteric strains of E. coli, or is immunospecific against SPATE of E. coli of the B2 or D or the B2+D genotype.
25 The immunoassay comprises the steps of: (i) contacting an enteric bacteria-containing sample from a subject with an antibody immunospecific against SPATE or an antibody immunospecific against Ag43 or both antibodies, under conditions suitable to form a complex between SPATE or Ag43 and
30 the antibody; and (ii) detecting presence or absence of the

complex. Presence of the complex indicates that the subject has or is susceptible to developing IBD.

The diagnostic or prognostic method may involve determining the level of SPATE or Ag43 in the sample, in which
5 case a higher level of SPATE or Ag43 compared to a cut-off value would indicate that the subject has IBD or or is susceptible to developing IBD.

We also describe a commercial package for diagnosing IBD or determining susceptibility to developing IBD in a
10 subject. The subject includes those having symptoms of IBD, or are suspected of having IBD, including UC or CD. The package comprises an agent for detecting serine protease autotransporter (SPATE), or an agent for detecting Ag43 in an enteric bacteria-containing sample from the subject, and
15 instructions for using the agent(s) to detect SPATE or Ag43 in the sample. The kit is therefore made to be used for diagnosing or prognosing IBD in the subject.

The agent, which is part of the commercial package, includes those agents useful for detecting SPATE or Ag43
20 nucleic acid. Such agents include those that detect a region of SPATE or a region or Ag43 which is conserved among enteric bacteria, a region of SPATE or a region or Ag43 which is conserved among enteric E. coli, a region of SPATE or a region or Ag43 which is specific to enteric E. coli, a region of SPATE
25 or a region or Ag43 which is conserved among virulent enteric strains of E. coli, a region of SPATE or a region or Ag43 which is specific to virulent enteric strains of E. coli, or a region of SPATE or a region or Ag43 which is specific to E. coli of the B2 or D or the B2+D genotype. The agent may be those used

to detect SPATE or Ag43 nucleic acid by polymerase chain reaction (PCR). In particular, the agent may comprise primer sequences designed to amplify and detect the target sequences.

The agent, which is part of the commercial package, also includes those agents useful for detecting SPATE or Ag43 polypeptide. For example, the agent may be for detecting SPATE or Ag43 polypeptide in an immunoassay. The agent may be an antibody which is immunospecific against SPATE or Ag43 of enteric bacteria, or against SPATE or Ag43 of enteric E. coli, or against SPATE or Ag43 of virulent enteric strains of E. coli, or against SPATE or Ag43 of E. coli of the B2 or D or the B2+D genotype. The commercial package may further comprise SPATE polypeptide or Ag43 polypeptide for generating a standard curve(s), as a standard against which the test level of SPATE polypeptide or Ag43 polypeptide is assessed as being statistically higher, lower, or equivalent, to normal.

We also describe a method of testing potential therapeutic agents for treating inflammatory bowel disease (IBD), the method comprising contacting a test compound with SPATE polypeptide or with Ag43 polypeptide and determining whether the test compound binds to SPATE or inhibits SPATE protease activity, or whether the test compound binds Ag43 or inhibits Ag43 aggregation activity or cell adhesion activity, wherein binding to SPATE or inhibition of SPATE protease activity, or binding to Ag43 or inhibition of Ag43 aggregation or cell adhesion indicates that the compound is a potential therapeutic agent for treating IBD.

We also describe a method for cultivating enteric microbes. The method comprises the steps of: a) resuscitating

an enteric bacteria-containing sample from a subject, including those having IBD or specifically UC or CD, by resuspension in buffered peptone water; and b) culturing the resuspension. The enteric bacteria being cultivated may be a wide range of
5 microbes including bacteria, or specific microbes such as E. coli. The enteric bacteria-containing sample from which enteric microbes are cultivated may be a colonoscopy tissue biopsy. In some embodiments, there is no need for washing the tissue or removing mucus before resuscitating the microbes in
10 the sample.

We also describe a method for phylogenetic identification of enteric bacteria associated with inflammatory bowel disease (IBD). The method comprises the steps of: (i) amplifying DNA from an enteric bacteria-containing sample from
15 a subject having IBD; (ii) amplifying DNA from an enteric bacteria-containing sample from a subject free of IBD; (iii) comparing the amplified DNA obtained from step (i) with amplified DNA obtained from step (ii) to mark out DNA that is associated with the subject having IBD; and (iv) performing
20 phylogenetic analysis to identify bacteria containing the DNA associated with the subject having IBD. In this method, the enteric bacteria-containing sample may be a colonoscopy tissue biopsy. The DNA amplification may comprise PCR amplification of a polymorphic region flanked by conserved sequences of
25 enteric bacteria, in which the PCR amplification employs primers specific for the conserved sequences. For example, the primers may be specific to at least a portion of the 16S rRNA gene, or to at least a portion of the 23S rRNA gene. The primers may be specific to at least a portion of both the 16S
30 and the 23S rRNA genes.

We also describe a method for phylogenetic identification of enteric bacteria associated with inflammatory bowel disease (IBD), but using DNA from cultivated bacteria. The method comprises the steps of: (i) amplifying DNA from enteric bacteria wherein the bacteria are from a subject having IBD and were cultivated according to the method described above; (ii) amplifying DNA from enteric bacteria wherein the bacteria are from an IBD-free subject and were cultivated according to the method described above; (iii) comparing the amplified DNA obtained from step (i) with amplified DNA obtained from step (ii) to mark out DNA from bacteria associated with the subject having IBD; and (iv) performing phylogenetic analysis to identify the bacteria associated with the subject having IBD.

15 BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a RISA analysis of biopsy samples from IBD and control tissues.

Figure 2 shows the proportions of *E. coli* and non-*E. coli* in biopsy tissues from controls, and UC and CD patients.

20 Figure 3A-D shows an alignment of SPATE nucleotide sequences (SEQ ID NOs:57-79) amplified using RISA on biopsy tissue.

Figure 4A-D shows an alignment of *pic*-like gene with other SPATE (SEQ ID NOs:1 and 57-79) genes.

25 Figure 5A-C shows an alignment of Ag43 nucleotide sequences (SEQ ID NOs:80-112). The Ag43 sequences shown here

encode, in the +3 reading frame, amino acids corresponding to amino acids 718-822 of SEQ ID NO:81.

DETAILED DESCRIPTION OF EMBODIMENTS

Our initial studies involve culture-independent surveys of microbial diversity. Our survey forms a prelude to targeted cultivation of bacteria, leading to our finding that SPATE and Ag43 are useful microbial markers of IBD. However, although our invention arose from our study and finds basis in our scientific results, the invention should not be limited to the specific results. This is understood to those working in the field of microbiology as it pertains to inflammatory bowel disease.

We note that, where microorganisms are identified using culture-independent methods as being associated with disease, there may be advantages to culturing them so that their virulence mechanisms can be evaluated.

The work described in the example here involve the use of ribosomal intergenic spacer analysis (RISA, ref 13) of biopsy tissue to identify bands that were consistently associated with IBD tissue. We used highly focused cultivation methods, including resuscitation methods, to specifically culture Enterobacteriaceae.

(I) SERINE PROTEASE AUTOTRANSPORTER (SPATE) and ANTIGEN 43 (AG43)

The term "SPATE" in the present context means nucleic acids and polypeptides which are members of the serine protease autotransporter family and which are homologs of the nucleotide

sequence identified as SEQ ID NO:1 and the amino acid sequence identified as SEQ ID NO:2. The term "Ag43" in the present context means nucleic acids and polypeptides which are members of the autotransporter family and which are homologs of the
5 nucleotide sequence identified as SEQ ID NO:80 and the amino acid sequence identified as SEQ ID NO:81. Autotransporter proteins including the SPATE and Ag43 family are described in ref 28.

Homologs of the SPATE family include, for example,
10 the SPATE sequences identified in the Genbank accession numbers: AF056581, AF218073, AF297061, AJ278144, AJ586888, AX276281, AY163491, AY258503, AY604009, DD002707, U69128, X97542, Y13614, NZ_AAJV01000028, Sat_AX702523. The homolog sequences also include those shown as SEQ ID NOS:1, 2
15 and 57-71.

We contemplate "SPATE" in the present invention to encompass naturally occurring homologs and sequence variants of SEQ ID NO:1 and 2. A SPATE homolog possesses the three domains that are typical of SPATE autotransporters: an unusually long
20 signal sequence of about 49 amino acids, a passenger domain containing a consensus serine protease active site (GDSGSP or GDSGSG); and a C-terminal autotransporter domain. "Homologous amino acid sequence" or "variant amino acid sequence" is any polypeptide which is encoded, in whole or in part, by a nucleic
25 acid sequence which hybridizes at 25-35°C below critical melting temperature (T_m), to any portion of the nucleic acid sequence of SEQ ID No: 1. A homologous amino acid sequence is one that differs from an amino acid sequence shown in SEQ ID No: 2 by one or more conservative amino acid substitutions.

Such a sequence encompasses those variants which retain at least one inherent characteristics of the polypeptide such as immunogenicity, serine protease activity, haemagglutinin activity, mucinase activity, elastase activity, cytotoxic effects on cells, elastase activity, lipoprotein cleavage activity, coagulation factor V cleavage activity, the ability to degrade the barrier function of the gut, and the ability to cleave proteins in the enterocyte. Such a sequence is contemplated as being at least 75%, 80%, 90% or 95% identical to SEQ ID No: 2. We contemplate a homolog or variant sequence to differ from the sequence of reference by a majority of conservative amino acid substitutions, i.e. substitutions among amino acids of the same class.

We contemplate "Ag43" in the present invention to encompass naturally occurring homologs and sequence variants of SEQ ID NO:80 and 81. "Homologous amino acid sequence" or "variant amino acid sequence" is any polypeptide which is encoded, in whole or in part, by a nucleic acid sequence which hybridizes at 25-35°C below critical melting temperature (T_m), to any portion of the nucleic acid sequence of SEQ ID No: 80. A homologous amino acid sequence is one that differs from an amino acid sequence shown in SEQ ID No: 81 by one or more conservative amino acid substitutions. Such a sequence encompasses those variants which retain at least one inherent characteristics of the polypeptide such as immunogenicity, auto-aggregation activity, cell to cell aggregation, the ability to induce a frizzy colony morphology and the ability to form a biofilm (see refs 27 and 39). Such a sequence is contemplated as being at least 75%, 80%, 90% or 95% identical to SEQ ID No: 81. We contemplate a homolog or variant sequence

to differ from the sequence of reference by a majority of conservative amino acid substitutions, i.e. substitutions among amino acids of the same class.

Homologs of the Ag43 family include, for example, the
5 Ag43 sequences identified in the Genbank accession numbers:
AE005174, AE014075, AF233271, AF233272, AJ303141, AJ586887,
AJ586888, AJ617685, AP009048, AR580480, AX370193, AX702425,
AX702524, AY857617, BA000007, BD184766, BD195283, BD444174,
CP000243, CS148067, U00096, U24429, X16664. The homolog
10 sequences also include those shown as SEQ ID NOS: 80-104.

Homology is measured using sequence analysis software
such as Sequence Analysis Software Package of the Genetics
Computer Group, University of Wisconsin Biotechnology Center,
1710 University Avenue, Madison, WI 53705. Amino acid
15 sequences are aligned to maximize identity. Gaps may be
artificially introduced into the sequence to attain proper
alignment.

In the present context, stringent conditions are
achieved for both pre-hybridizing and hybridizing incubations
20 (i) within 4-16 hours at 42°C, in 6 x SSC containing
50% formamide, or (ii) within 4-16 hours at 65°C in an aqueous
6 x SSC solution (1 M NaCl, 0.1 M sodium citrate (pH 7.0)).
Typically, hybridization experiments are performed at a
temperature from 60 to 68°C, e.g. 65°C. At such a temperature,
25 stringent hybridization conditions can be achieved in 6xSSC,
preferably in 2xSSC or 1xSSC, more preferably in 0.5xSSC,
0.3xSSC or 0.1xSSC (in the absence of formamide). 1xSSC
contains 0.15 M NaCl and 0.015 M sodium citrate.

In particular, we contemplate SPATE sequences which contain the conserved serine protease motif GDSGSP or GDSGSG (corresponding to amino acids 195-200 of SEQ ID NO:2).

In the context of diagnosis or prognosis or
5 evaluating the effectiveness of a treatment for IBD, i.e. in the context where SPATE or Ag43 is to be detected in a sample from a subject, we contemplate the naturally occurring SPATE or Ag43 sequences, examples for which are set out above.

In the context of testing potential therapeutic
10 agents for treating IBD, we contemplate using not only SPATE or Ag43 polypeptides and nucleic acids having naturally-occurring sequences, but also SPATE or Ag43 fragments; in particular fragments containing the conserved serine protease motif GDSGSP or GDSGSG.

15 The term "isolated polynucleotide or nucleic acid or polypeptide" is defined as a polynucleotide or nucleic acid or polypeptide removed from the environment in which it naturally occurs. For example, a naturally-occurring DNA molecule present in the genome of a living bacteria or as part of a gene
20 bank is not isolated, but the same molecule separated from the remaining part of the bacterial genome, as a result of, e.g., a cloning event (amplification), is isolated. Typically, an isolated DNA molecule is free from DNA regions (e.g., coding regions) with which it is immediately contiguous at the 5' or
25 3' end, in the naturally occurring genome. Such isolated polynucleotides may be part of a vector or a composition and still be defined as isolated in that such a vector or composition is not part of the natural environment of such polynucleotide.

The polynucleotide for use in certain aspects of the invention is either RNA or DNA (cDNA, genomic DNA, or synthetic DNA), or modifications, variants, homologs or fragments thereof. The DNA is either double-stranded or single-stranded, and, if single-stranded, is either the coding strand or the non-coding (anti-sense) strand. By "polypeptide" or "protein" is meant any chain of amino acids, regardless of length or post-translational modification (e.g., glycosylation or phosphorylation). Both terms are used interchangeably in the present application. Polynucleotide and nucleic acid are interchangeable terms as used herein.

Partial sequences of SPATE or Ag43, e.g. of SEQ ID No:2 or 81, or their homologous amino acid sequences, are inherent to the full-length sequences. Such polypeptide fragments preferably are at least 12 amino acids in length, preferably at least 15, 20, 25, 30, 35, 40, 45, 50 amino acids, more preferably at least 55, 60, 65, 70, 75 amino acids, and most preferably at least 80, 85, 90, 95, 100 amino acids in length.

In the present context, fusion polypeptides may be useful, for example for testing potential therapeutic agents that may bind or inhibit SPATE or Ag43. A fusion polypeptide is one that contains a polypeptide or a polypeptide derivative of the invention fused at the N- or C-terminal end to any other polypeptide (hereinafter referred to as a peptide tail). A simple way to obtain such a fusion polypeptide is by translation of an in-frame fusion of the polynucleotide sequences, i.e., a hybrid gene.

(II) Assays for SPATE and Ag43 nucleic acids

The assay for SPATE and Ag43 can involve direct assay of nucleic acid levels, such as mRNA levels to measure gene expression, or involve direct assay of DNA, for example by PCR, to gauge the number of gene copies, thereby estimating the level of SPATE- or Ag43-containing, and possibly pathogenic, microbes.

The SPATE or Ag43 nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of SPATE or Ag43 nucleic acid and SPATE- or Ag43-containing microbes. Experimental data as provided herein indicates that SPATE or Ag43 and virulent enterobacteria expressing SPATE or Ag43 are associated with IBD. Accordingly, probes based on SPATE or Ag43 sequences can be used to detect the presence of, or to determine levels of, SPATE or Ag43 in cells, tissues, and in organisms of the gut. The nucleic acid whose level is being determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in SPATE or Ag43 expression, or involving an increase or decrease in SPATE or Ag43 DNA-containing microbes, relative to normal results.

Nucleic acids can be detected by methods known in the art. RNA may be detected by for example Northern analysis or by the reverse transcriptase-polymerase chain reaction (RT-PCR) method (see for example Sambrook et al (1989) Molecular Cloning: A Laboratory Manual (second edition), Cold Spring

Harbor Laboratory Press, Cold Spring Harbor, New York, USA).

DNA is routinely detectable by PCR, Southern hybridizations and in situ hybridization.

Probes can be used in a diagnostic test for
5 identifying samples containing SPATE or Ag43 or SPATE- or Ag43-expressing microbes, such as by measuring a level of nucleic acid encoding SPATE or Ag43 in a sample of microbes from a subject.

Detection of SPATE or Ag43 nucleic acids may be
10 designed such that specific sub-families of SPATE or Ag43 are detected, e.g. those of enteric bacteria, enteric E. coli, or virulent enteric strains of E. coli. It is also useful to detect SPATE or Ag43 nucleic acids specific to certain sub-families, e.g. SPATE or Ag43 specific to enteric bacteria,
15 enteric E. coli, virulent enteric strains of E. coli, or E. coli of the B2 or D genotype or the B2+D genotype. By choosing the regions of SPATE or Ag43 sequences which are conserved or unique to the various sub-families, one can target detection of certain SPATEs or Ag43s. For example, a probe or PCR primers
20 complementary to a SPATE or Ag43 sequence that is conserved among enteric E. coli can be used to detect the presence of SPATE or Ag43, and thereby the presence of enteric E. coli, in the sample, which is diagnostic or prognostic of IBD.

(III) ASSAYS FOR SPATE OR AG43 POLYPEPTIDE

25 SPATE or Ag43 proteins are useful targets for diagnosing IBD or predisposition to IBD. The invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a sample of cell, tissue or body

fluid containing enteric bacteria. Experimental data as provided herein shows SPATE or Ag43 to be a microbial indicator of IBD. Thus to diagnosis or prognosis of IBD involves contacting an enteric microbe-containing sample with a compound
5 capable of interacting with SPATE or Ag43, such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

SPATE protein levels may be detected either directly
10 using affinity reagents, e.g. an antibody or fragment thereof (for methods, see for example Harlow, E. and Lane, D (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY)) or by assaying the protein's activity such as serine protease activity (ref 38),
15 haemagglutinin activity, mucinase activity, elastase activity (ref 29 & 30), cytotoxicity (ref 31-33), elastase activity (ref 27-30), lipoprotein cleavage activity (ref 27-30), coagulation factor V cleavage activity.

Ag43 protein levels may be detected either directly
20 using affinity reagents, e.g. an antibody or fragment thereof (for methods, see for example Harlow, E. and Lane, above) or by assaying the protein's activity such as auto-aggregation activity, cell to cell aggregation, the ability to induce a frizzy colony morphology and the ability to form a biofilm (see
25 refs 27 and 39).

One useful agent for detecting SPATE or Ag43 protein in a sample is an antibody capable of selectively binding to SPATE or Ag43. Binding may be selective with respect to specific sub-families of SPATE or Ag43, e.g. those of enteric

bacteria, enteric E. coli, or virulent enteric strains of E. coli, or E. coli of the B2 or D or the B2+D genotype. By choosing antibodies immunospecific for the regions of SPATE or Ag43 which are conserved or unique to the various sub-families, one can target detection of certain SPATEs or Ag43s. For example, an antibody immunospecific for a SPATE or Ag43 epitope that is a conserved sequence among enteric E. coli can be used to detect enteric E. coli in the sample, which is diagnostic or prognostic of IBD.

For detecting SPATE or Ag43 protein and thereby diagnosing or prognosing IBD in a subject, an antibody may be use in a method that includes contacting an enteric bacteria-containing sample from a subject with an anti-SPATE or anti-Ag43 antibody, under conditions suitable to form a complex between SPATE or Ag43 and the antibody; and detecting the presence or absence of the complex. The presence of the complex indicates that the subject has or is susceptible to developing IBD. The presence or absence of the complex can be detected, for example, with a detectable secondary antibody that has specificity for a class determining portion of the primary antibody. The term "complex" is used synonymously here with "immune complex" and means an aggregate of two or more molecules that results from specific binding between an antigen (SPATE or Ag43) and an antibody.

An antibody of the invention is either polyclonal or monoclonal. Monospecific antibodies may be recombinant, e.g., chimeric (e.g., constituted by a variable region of murine origin associated with a human constant region), humanized (a human immunoglobulin constant backbone together with hypervariable region of animal, e.g., murine, origin), and/or

single chain. Both polyclonal and monospecific antibodies may also be in the form of immunoglobulin fragments, e.g., F(ab)'₂ or Fab fragments. The antibodies of the invention are of any isotype, e.g., IgG or IgA, and polyclonal antibodies are of a
5 single isotype or a mixture of isotypes.

Antibodies against the polypeptides, homologs or fragments of the present invention are generated by immunization of a mammal with a composition comprising said polypeptide, homolog or fragment. Such antibodies may be
10 polyclonal or monoclonal. Methods to produce polyclonal or monoclonal antibodies are well known in the art. For a review, see "Antibodies, A Laboratory Manual, Cold Spring Harbor Laboratory, Eds. E. Harlow and D. Lane (1988), and D.E. Yelton et al., 1981. Ann. Rev. Biochem. 50:657-680. For monoclonal
15 antibodies, see Kohler & Milstein (1975) Nature 256:495-497.

The antibodies of the invention, which are raised to a polypeptide or polypeptide derivative of the invention, are produced and identified using standard immunological assays, e.g., Western blot analysis, dot blot assay, or ELISA (see,
20 e.g., Coligan et al., Current Protocols in Immunology (1994) John Wiley & Sons, Inc., New York, NY). The antibodies are used in diagnostic methods to detect the presence of a SPATE or At43 antigen in a sample, such as a biological sample. The antibodies are also used in affinity chromatography for
25 purifying a polypeptide or polypeptide derivative of the invention.

Those skilled in the art will readily understand that the immune complex is formed between a component of the sample and the antibody, polypeptide, or polypeptide derivative,

whichever is used, and that any unbound material is removed prior to detecting the complex. It is understood that a polypeptide reagent is useful for detecting the presence of anti-SPATE or anti-Ag43 antibodies in a sample, while an anti-
5 SPATE or anti-Ag43 antibody is useful for screening a sample, such as a gastric extract or biopsy, for the presence of SPATE or Ag43 polypeptides.

A secondary antibody can be, for example, an anti-IgA secondary antibody, an anti-IgG secondary antibody, or a
10 combination of anti-IgA and anti-IgG secondary antibodies.

In vitro techniques for detection of SPATE or Ag43 include enzyme linked immunosorbent assays (ELISAs), Western blots, immuno-precipitations and immunofluorescence using a detection reagent, such as an antibody or protein binding
15 agent. Alternatively, the protein can be detected in vivo in a subject by introducing into the subject a labeled antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging
20 techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not
25 significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target such that there are

shared epitopes. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms
5 consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The term "antibody" as used herein include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not
10 limited to, Fab or F(ab')₂, and Fv fragments.

Detection of an antibody of the present invention, and thus detection of an antibody-SPATE or antibody-Ag43 complex, can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of
15 detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, beta-galactosidase, or acetylcholinesterase;
20 examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a
25 luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ¹²⁵I, ¹³¹I, ³⁵S or ³H.

The antibodies can be used to isolate SPATE or Ag43 or fragments thereof by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the protein from natural sources and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Such antibodies can also determine the pattern of colonisation of SPATE-expressing microbes or Ag43-expressing microbes along the gastro-intestinal tract.

Experimental data as provided herein indicates that SPATE or Ag43 and virulent enterobacteria expressing SPATE or Ag43 are associated with IBD. Antibodies to SPATE or Ag43 can be used to detect SPATE or Ag43 protein in situ, in vitro, or in a lysate or supernatant of an enteric bacteria-containing sample in order to evaluate the abundance and pattern of expression. The antibodies can also be used to assess a predisposition toward IBD, i.e. a prognosis of IBD.

(IV) SPATE AND IBD; AG43 AND IBD

SPATEs are potentially important in IBD because they exhibit functions like degradation of the barrier function of the gut, and cleavage of proteins in the enterocyte, all of which are phenotypes associated with IBD (ref 28). For example Vat, Pic, and Pic-like have haemagglutinin, mucinase, and elastase activity (ref 29, 30), Sat has cytotoxic effects (ref 31-33) on cells as well as elastase activity (ref 27-30), and EspP cleaves lipoproteins (ref 27-30). Functional studies on

SPATE have demonstrated that Pic and Sat can cleave coagulation factor V, potentially linking it to haemorrhagic events in the gut. Additionally, Pic is thought to be involved in colonisation of *E. coli* to intestinal tissue (ref 34). Ag43 is a surface
5 adhesin that promotes bacterial biofilm formation due to cell-to-cell aggregation (ref 27).

SPATE or Ag43 can thus be used as a diagnostic marker for IBD, or to determine susceptibility to developing IBD in a subject. The presence of SPATE or Ag43 in an enteric bacteria-
10 containing sample from a subject compared to a control indicates that the subject has IBD or is susceptible to developing IBD.

As used herein, the term "subject" means any animal capable of having inflammatory bowel disease, including a
15 human, non-human primate, rabbit, rat or mouse, especially a human. A subject can have one or more symptoms of Crohn's disease or ulcerative colitis, or may be asymptomatic. The term "subject having IBD" means a subject having the clinical features of IBD as defined herein. The term "susceptible to
20 IBD" as indicated by the presence of SPATE or Ag43 in the gut micro flora of the subject means a reduced ability to resist IBD-causing factors, as compared with an individual from whom a sample is obtained that does not contain a significant level of SPATE or Ag43 or SPATE- or Ag43-expressing microbes.
25 Susceptibility to IBD in a subject does not mean the subject will develop IBD, but that the subject has an increased probability of having symptoms of IBD in the future.

Inflammatory bowel disease (IBD) encompasses a group of diseases such as ulcerative colitis (UC) and Crohn's disease

(CD). IBDs can be difficult to diagnose. An initial diagnosis, made on the basis of medical history and physical examination, is generally confirmed via imaging tests to look at the intestines and laboratory culture tests to rule out bacterial, viral and parasitic infections. Crohn's disease affects some areas of the intestines and not others. Ulcerative colitis is more dispersed. Endoscopy is used to take a biopsy of intestinal tissue, which can be used to identify the deep inflammation of the bowel that is characteristic of Crohn's disease. X-rays (after oral or rectal ingestion of Barium), computed tomography (CT) scan, and magnetic resonance imaging (MRI) may be helpful in locating fistulas. A stool analysis (including a test for blood in the stool) is often performed, depending on symptoms, to look for blood and signs of bacterial infection. Blood and urine tests may be done to check for anemia, high white cell counts, or malnutrition; all these are signs of IBDs.

Murine models of inflammatory bowel disease can be used in certain aspects of the invention, e.g. to evaluate a potential treatment for IBD using a compound identified by screening with SPATE or Ag43. Animal models are known to those in the art, for example, mice with targeted disruption of the gene encoding the alpha subunit of the G-protein Gi2 exhibit features of human bowel disease. Mice deficient in IL-10 and mice deficient in IL-2 also have colitis-like disease.

A suitable sample is any sample containing enteric bacteria. Sources for samples include gut tissue biopsy, intestinal mucosa, stool or fecal matter, or intestinal wash. In particular, the sample may be a colonoscopy tissue biopsy from the lower gastrointestinal (GI) tract. Biopsies resected

from the gastro-intestinal tract and from an area believed to be exhibiting signs of the disease may also be useful as the sample source. The sample may be used directly in the methods described herein. Alternatively, the sample may be processed
5 e.g. to remove particulate matter or to remove mucus etc. as appropriate for a chosen technique.

In the methods described herein, one may determine the level of SPATE or Ag43 in the sample, rather than simply detecting the presence of SPATE or Ag43 compared to the absence
10 of SPATE or Ag43. A higher level of SPATE or Ag43 compared to a cut-off value would indicate that the subject has IBD or is susceptible to developing IBD. By "higher level", we mean a quantitative rather than qualitative difference since "absence" or "presence" are relative terms, a test sample result being
15 always to be compared with a control. An appropriate control are samples from subjects free of IBD symptoms, or possibly samples from the same IBD subject but obtained from a region of the GI tract that contains enteric bacteria but is free of inflammation or any sign of IBD. To normalize the SPATE or
20 Ag43 values obtained from the assays, one may assay for a protein or gene which is known to be present equivalently in both IBD and non-IBD subjects. The SPATE or Ag43 values may be normalized against such control values. Since SPATE or Ag43 may be present in non-pathogenic bacteria, there may be a
25 baseline level of SPATE or Ag43 in non-IBD subjects. Such a baseline level establishes a cut-off value for determining whether a subject has more-than-normal SPATE or Ag43.

The invention also encompasses commercial packages or kits for diagnosing IBD or determining susceptibility to
30 developing IBD in a subject. The package comprises an agent

for detecting SPATE or Ag43 in an enteric bacteria-containing sample from the subject. The package may also contain instructions for using the agent to detect SPATE or Ag43 in the sample, thereby diagnosing or prognosing IBD in the subject.

5 If SPATE or Ag43 nucleic acid is to be detected, the agent may be a nucleic acid probe or a set of primers for use in PCR amplification. If SPATE or Ag43 protein is to be detected, the agent may be an antibody immunospecific for SPATE or Ag43; the package may also contain secondary antibodies to detect the
10 SPATE-antibody or Ag43-antibody complex. The package may also contain SPATE or Ag43 DNA or protein or cells expressing SPATE or Ag43, in unit amounts suitable as standards against which the test results are assessed. The package may also contain reagents or materials for detecting a normalizing protein or
15 gene (one which is known to be present equivalently in both IBD and non-IBD subjects) against which the SPATE or Ag43 test results are assessed.

(V) SCREENING ASSAYS FOR POTENTIAL THERAPY

Nucleic acid expression assays are useful for drug
20 screening to identify compounds that modulate SPATE or Ag43 gene expression or modulate growth of SPATE- or Ag43-expressing microbes. The invention thus provides a method for identifying a compound that can be used to treat a disorder of the gastrointestinal tract associated with expression of SPATE or Ag43,
25 particularly biological and pathological processes that involve SPATE or Ag43. The method typically includes assaying the ability of the compound to modulate expression of the pertinent SPATE or Ag43 gene and thus identify a compound that can be used to treat a disorder characterized by SPATE or Ag43 gene
30 expression.

Thus SPATE or Ag43 may be used as a target in screening assays to identify compounds that are useful as inhibitors of SPATE or Ag43 for the prevention or treatment of IBD. In some embodiments, such an assay may comprise the steps
5 of: (a) providing a test compound; (b) providing a source of SPATE or Ag43; and (c) measuring SPATE or Ag43 activity in the presence versus the absence of the test compound. A lower measured activity in the presence of the test compound would indicate that the compound is an inhibitor of SPATE- or Ag43-
10 dependent activity and may be useful for the prevention and/or treatment of IBD.

"SPATE activity" as used herein refers to any type of observed phenomenon which can be attributed to SPATE. Such activity includes serine protease activity (ref 38),
15 haemagglutinin activity, mucinase activity, elastase activity (ref 29 & 30), cytotoxicity (ref 31-33), elastase activity (ref 27-30), lipoprotein cleavage activity (ref 27-30), coagulation factor V cleavage activity. Another inherent activity that can be assayed for is immunogenicity. A test compound that can
20 mitigate or block a SPATE-specific epitope may be effective.

"Ag43 activity" as used herein refers to any type of observed phenomenon which can be attributed to Ag43. Such activity includes auto-aggregation activity, cell to cell aggregation, the ability to induce a frizzy colony morphology
25 and the ability to form a biofilm (see refs 27 and 39). Another inherent activity that can be assayed for is immunogenicity. A test compound that can mitigate or block a Ag43-specific epitope may be effective.

The assay may be carried out *in vitro* utilizing a source of SPATE or Ag43 which may comprise naturally isolated or recombinantly produced SPATE or Ag43, in preparations ranging from crude to pure. Such assays may be performed in an
5 array format.

The assay may in an embodiment be performed using an appropriate host cell as a source of SPATE or Ag43. Such a host cell may be prepared by the introduction of DNA encoding SPATE or Ag43 into the host cell and providing conditions for
10 the expression of SPATE or Ag43.

SPATE or Ag43 and fragments, particularly those SPATE fragments comprising the conserved protease motif GDSGSP or GDSGSG, are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e.
15 microbes that normally express SPATE or Ag43. In an alternate embodiment, cell-based assays involve recombinant host cells expressing SPATE or Ag43 or their fragments.

SPATE or Ag43 can be used to identify compounds that modulate its role in IBD. SPATE or Ag43 and appropriate
20 variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to SPATE or Ag43 and their fragments. These compounds can be further screened against a functional SPATE or Ag43 to determine the effect of the compound on SPATE or Ag43 activity. Further,
25 these compounds can be tested in animal systems to determine activity/effectiveness.

Binding and/or activating compounds can also be screened by using fusion proteins in which the amino terminal

domain, or parts thereof, and the carboxy terminal domain, or parts thereof, can be replaced by heterologous polypeptides. These are generally referred to as chimeric or fusion proteins.

SPATE or Ag43 and fragments, particularly those SPATE
5 fragments comprising the conserved protease motif GDSGSP or GDSGSG, are also useful in competition binding assays in methods designed to discover compounds that interact with SPATE or Ag43 (e.g. binding partners and/or ligands). Thus, a compound is exposed to SPATE or Ag43 and fragments under
10 conditions that allow the compound to bind or to otherwise interact with the polypeptide. A known binding partner such as a monoclonal antibody to SPATE or to Ag43 is also added to the mixture. If the test compound interacts with SPATE or Ag43, it may decrease the amount of complex formed between SPATE or Ag43
15 and the known binding partner. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of SPATE or Ag43. Thus, the binding partner that competes with the test compound is designed to bind to peptide sequences corresponding to the
20 region of interest in SPATE or Ag43.

The above-described assay methods may further comprise determining whether any compound so identified can be used for the prevention or treatment of IBD, such as examining their effect(s) on disease symptoms in suitable animal model
25 systems. Furthermore, one can examine their effect(s) on the SPATE- or Ag43-expressing microbe to determine whether the compound adversely affects pathogenic microbes, e.g. inhibiting their growth and functioning as a microbe-specific antibiotic.

(VI) CULTIVATING ENTERIC BACTERIA AND PHYLOGENETIC
IDENTIFICATION OF IBD-ASSOCIATED MICROBES

We found that we could culture bacteria from untreated biopsy tissue by resuscitating the sample in buffered peptone water (a media used for the preliminary, non-selective enrichment of bacteria, particularly pathogenic Enterobacteriaceae, from foodstuffs and other materials; available for example at Genpharm Inc. 85 Advance Road, Etobicoke, Ontario M8Z 2S9 Canada).

10 We further worked on a culture-independent strategy for picking up bacteria that are present preferentially on IBD tissue. We applied nucleic acid-based techniques to identify DNA segments more commonly present in IBD than in controls. Our method involves: (i) amplifying DNA from an enteric
15 bacteria-containing sample from a subject having IBD; (ii) amplifying DNA from an enteric bacteria-containing sample from a subject free of IBD; (iii) comparing the amplified DNA obtained from step (i) with amplified DNA obtained from step (ii) to mark out DNA that is associated with the subject having
20 IBD; and (iv) performing phylogenetic analysis to identify bacteria containing the DNA associated with the subject having IBD.

The overall goal of our strategy is to obtain as diverse a survey of the micro flora as possible, and identify
25 from the results those associated with IBD. To obtain a diverse cross section of the micro flora, the method involves obtaining microbial nucleic acids in regions that are variable, in order that the polymorphisms can point to the phylogeny of the microbe. One way to do this involves targeting a region

flanked by conserved sequences of enteric bacteria. The polymorphic analysis can be performed directly on the sample, or preferably on extracted DNA. Suitable DNA extraction techniques for extracting the total DNA from the various type
5 of samples are well known, and the appropriate method would easily be determined.

The DNA can be amplified by known methods, such as for example, the PCR method. In PCR, preferably a primer would be directed towards a conserved region to ensure that the
10 largest population of micro flora DNA is amplified, while the area amplified includes a less conserved region, thereby allowing a broad polymorphic analysis. Suitable examples might, for example, include the 16S rRNA gene, 23S rRNA gene or the region between the 16S and 23S rRNA genes. Any form of
15 polymorphic analysis is suitable. The more variable products that are detectable, the more determinate the analysis will be. For example, a restriction fragment length polymorphism analysis could be performed over the variable region of the 16S rRNA gene.

20 Comparison of the polymorphic analysis of the DNA from the sample of unknown origin to the DNA of the sample of known origin can be achieved using a variety of methods. For example, for a restriction fragment length polymorphism analysis, comparison may be achieved by a visual comparison of
25 an autoradiograph of a polyacrylamide gel electrophoresis, or alternatively the PCR product could be fluorescently-tagged and then laser detected and the electropherogram may be visually compared. Alternatively, the polymorphic profiles may be compared mathematically. A suitable method for determining E.
30 coli phylogenetic group is described in ref 15.

(VII) EXPERIMENTAL BASIS:

MATERIALS AND METHODS

Study subjects: We utilized 84 biopsies from 15 controls, 13 Crohn's disease (CD) patients (3 with ileal disease, 6 with ileocolonic disease and 4 with isolated colonic disease), and 19 ulcerative colitis (UC) patients (3 with proctitis, 8 with left sided colitis, and 8 with pancolitis) from a population-based case-control study undertaken at the University of Manitoba (table 1) as described (ref 14).

10 In brief, a population-based study refers to a process by which selection of study subjects proceeds by accounting for bases related to various factors like lifestyle (e.g. smoking), geographic location (e.g. urban vs rural), age, gender, or ethnicity. In IBD research it is challenging to
15 obtain untainted biopsy controls because endoscopy is normally only performed on persons when it is clinically required. The controls were true controls in the sense that the subjects voluntarily submitted to endoscopy and were drawn from the same population-based study. No antibiotics were provided to any of
20 the subjects in the six weeks prior to the colonoscopy.

Table 1 Biopsy samples used in this study*

	Controls (15 subjects 28 tissues)		CD (13 patients 27 tissues)		UC (19 patients 29 tissues)		IBD (32 patients 56 tissues)	
Site	End.	Hist.	End.	Hist.	End.	Hist.	End.	Hist.
Rectum	0 (13)	0 (13)	5 (13)	5 (13)	9 (10)	9 (10)	14 (23)	14 (23)
Caecum	0 (15)	0 (15)	4 (9)	6 (9)	2 (15)	4 (15)	6 (24)	10 (24)
Colon	0 (0)	0 (0)	3 (5)	3 (5)	2 (4)	2 (4)	5 (9)	5 (9)

*Endoscopic (End.) and histological (Hist.) examination was made on biopsies and only histologically positive samples were considered inflamed. Total numbers of biopsies for each group are presented in parentheses, and the numbers not in parentheses represent inflamed biopsies.

Colonoscopy plus biopsies: Following a standard oral Fleet® Phospho-soda® treatment, biopsies were taken from the caecum
5 the rectum. In subjects with a previous caecal resection, biopsies were obtained from the right colon distal to the ileocolonic anastmosis. All biopsies were snap frozen in liquid nitrogen and stored at -70°C. Biopsies were subject to standard
10 histological staining with haematoxylin and eosin for evaluation of inflammation. A site was considered inflamed if it had histological evidence of inflammation and was considered uninflamed if it was histologically normal.

DNA extraction for RISA analysis: Tissue samples were suspended in 150 µl lysis buffer [10 mM Tris-HCl, pH 8.0; 5 mM
15 EDTA, pH 8.0; 4 M guanidinium isothiocyanate (GITC), pH 7.5; 50 g Sarcosyl/L, 2.5 g SDS/L, 5 g sodium citrate/L and 5 g Triton X-100/L]. 300 µl of chloroform and Tris-saturated phenol (pH 6.9) were added to each tube. The samples were placed

at -20°C for 1 h. Subsequently, samples were centrifuged in microfuge tubes at 4°C for 20 min at 10,000 x g. Supernatants were transferred to fresh tubes. Isopropanol to 1/4 volume of the supernatants was added and the mixtures loaded onto silica-cellulose membranes in columns. Samples were allowed to filter through the membrane by gravity. The membranes were washed twice with 300 µl 95% ethanol (by gravity). DNA was eluted with 400 µl hot (about 75°C) TE buffer (by gravity) and precipitated with two portions of 95% ethanol. The resulting pellets were suspended in 25 µl 0.5x TE buffer (pH 8.0) and stored at -20°C until further analysis.

Primer sequences used for RISA are listed in table 2 and amplified intergenic transcribed spacers between the 16S and 23S rDNA13. PCR products were subjected to electrophoresis using 2% agarose. DNA fragments only found in UC and CD were purified from agarose gels, cloned into the pCR®2.1-TOPO TA vector (Invitrogen) and sequenced. Standard bioinformatics analysis was used to taxonomically classify the sequence fragment.

Bacterial cultures: Once RISA analysis had determined that the bacteria that appeared in UC and CD and not in controls were E. coli, targeted bacterial cultivation was carried out. All cultivation of the bacteria was with untreated biopsy tissue and no procedures were used to wash the tissues or to remove mucus. To ensure as many E. coli cells as possible were cultured, resuscitation in buffered peptone water was performed, followed by decimal dilution and culturing on chromogenic E. coli / coliform medium (Oxoid CM0956). Resuscitation was by incubation of the biopsy in 1 ml of 100 mM buffered peptone water for 16 hours at 37°C. Ten µl droplets

were pipetted (Maxipettor, Eppendorf) onto media, allowed to dry, and then inverted and incubated at 37°C. After 18 hours, *E. coli* (purple colonies) and non-*E. coli* coliforms (blue, pink, and white colonies) were counted separately.

5 Colour differentiation on chromogenic agar is a good first approximation of *E. coli* identity. Five putative *E. coli* (purple colonies) were picked from each positive tissue and sub-cultured in LB broth, then recultured on *E. coli* / coliform medium, and tested for reactivity to indole, methyl red, Vogues
10 Proskauer, and citrate utilisation to differentiate *E. coli* from non-*E. coli*. The 16S rDNA gene sequences for all the non-*E. coli* were determined using standard primers (table 2).

DNA extraction from bacterial cultures: For DNA extraction from cultures, 1 ml suspensions of each culture were
15 centrifuged and pellets were suspended in lysis buffer and subsequently mixed with chloroform and Tris-saturated phenol. All other steps were the same as for the extraction of DNA from tissue samples.

Molecular analysis of bacterial cultures: All primer sequences
20 used for molecular analysis of bacterial cultures are listed in table 2. Primers were either modified from published primers, or newly designed for the purposes of our studies. PCR was by standard methods after optimisation for specificity of primers pairs by adjusting annealing temperature and salt
25 concentration. Amplified products were run on agarose gels as described above.

Table 2 Primers used			
Primer sequences	Target sequences	SEQ ID NO	
Autotransporters			
SPATE1 5' GAGGTCAACAACCTGAACAAACGTATGGG	The genes encoding serine protease autotransporters (SPATE)	3	
SPATE2 5' CCGGCACGGGCTGTCACTTTCCAG		4	
E. coli phylogenetic groups			
ChuAf 5' CGGACGAACCAACGGTCAGGAT	The chuA gene is required for heme transport in E. coli O157:H7	5	
ChuAr 5' TGCCGCCAGTACCAAAGACACG		6	
Yjaf 5' CGTGAAGTGTGAGGAGACGCTGC	The yjaA gene coding for protein of unknown function	7	
Yjar 5' TCGGTTCTCAACCTGTGACAAACC		8	
Tsp1 5' GGGAGTAATGTCGGGGCATTCAG	Tsp encodes for a putative DNA fragment (TSPE4.C2) in E. coli	9	
Tsp2 5' CATCGCGCCAACAAAGTATTACGCAG		10	
E. coli toxins			
Cnff 5' AGTACTGACACTCACTCAAGCCGC	Cytotoxic necrotising factors (Cnf1 and Cnf2)	11	
Cnfr 5' GCAGAACGACGTTCTTCATAAGTATCACC		12	
IpgDf 5' CGACTTCTCTTCTGACGCCGAC	ipgD gene modulates entry of bacteria into epithelial cells	13	
IpgDr 5' CAACATTCCTCCAGCCTAAGCCC		14	
Vt1f 5' CGCATAGTGGAACCTCACTGACGC	Verocytotoxin 1	15	
Vt1r 5' CATCCCCGTACGACTGATCCC		16	
Vt2f 5' CGGAATGCAAATCAGTCGTCCTCAC	Verocytotoxin 2	17	
Vt2r 5' TCCCCGATACTCCGGAAGCAC		18	
HlyAf 5' TGCAGCCTCCAGTGCATCCCTC	The hlyA gene encoding alpha hemolysin	19	
HlyAr 5' CTTACCACTCTGACTGCGATCAGC		20	
STaf 5' GTGAAACAACATGACGGGAGG	Heat-stable enterotoxin 1	21	
STar 5' ATAACATCCAGCACAGGCAGG		22	
STbf 5' GGGGTTAGAGATGGTACTGCTGGAG	Heat-stable enterotoxin 2	23	
STbr 5' GACAATGTCCGTCTTGCGTTAGGAC		24	
LTF 5' CCGTGCTGACTCTAGACCCCCA	Heat-labile enterotoxin LT	25	
LTr 5' CCTGCTAATCTGTAACCATCCTCTGC		26	
Eaef 5' CCAGGCTTCGTACAGTTGCAGGC	The eae gene coding for inti-min present in AEEC strains	27	
Eaer 5' CGCCAGTATTCGCCACCAATACC		28	
Enterotoxigenic Bacteroides fragilis (ETBF)			
Bf1F 5' GTTAGTGCCCAGATGCAGGATGCGG	The genes coding for	29	

Bf2F 5' GAACTCGGTTTATGCAGTTCATGGACTG	Bacteroides fragilis	30
Bf3R 5' TGGGTGTAGACATCCCACTGGCTT	enterotoxins (BFT1, Bft2 and	31
Bf4R 5' GGATACATCAGCTGGGTGTAGACATCCC	BFT3)	32
E. coli adhesins		
PapF 5' CCGGCGTTCAGGCTGTAGCTG	The genes coding for	33
PapR 5' GCTACAGTGGCAGTATGAGTAATGACCGTTA	pathogenicity islands (PAI I,	34
PAI1 5' TAGCTCAGACGCCAGGATTTTCCCTG	PAI II) ¹⁶ and sfp gene cluster	35
PAI2 5' CCTGGCGCCTGCGGGCTGACTATCAGGG		36
BmaEf 5' CTAACCTGCCATGCTGTGACAGTA	The bmaE gene for M-	37
BmaEr 5' TTATCCCCTGCGTAGTTGTGAATC	agglutinin subunit; Afa-8	38
	gene cluster	
Sfaf 5' CGGAGGAGTAATTACAAACCTGGCA	S-fimbrial adhesins encoded	39
Sfar 5' CTCCGGAGAACTGGGTGCATCTTAC	by sfaD to sfaE	40
Afaf 5' TATGGTGAGTTGGCGGGGATGTACAGTTACA	AfaE-3 gene cluster	41
Afar 5' CCGGGAAAGTTGTCGGATCCAGTGT		42
AIDA1 5' TATGCCACCTGGTATGCCGATGAC	The aidA gene coding E. coli AIDA-I	43
AIDA2 5' ACGCCCACATTCCCCCAGAC	adhesin in DAEC strains	44
AggRf 5' GAGTTAGGTCACTCTAACGCAGAGTTG	The aggR gene for adhesin of	45
AggRr 5' GACCAATTCGGACAAGTCAAGCATCTAC	aggregative adherence fimbria	46
	I	
Ag43F 5' TGACACAGGCAATGGACTATGACCG	The agn43 gene coding for	47
Ag43R 5' GGCATCATCCCGGACCGTGC	antigen involved in E. coli	48
	autoaggregation	
Flagella and type 1 fimbriae typing		
Primer1 5' CAAGTCATTAATAC(A/C)AACAGCC	The fliC genes coding for	49
Primer2	flagellin proteins	
5' GACAT(A/G)TT(A/G)GA(G/A/C)ACTTC(G/C)GT		50
FimHf 5' CTGGTCATTGCGCTGTAAAACCGCCA	The fimH gene encoding FimH	51
FimHr 5' GTCACGCCAATAATCGATTGCACATTCCCT	subunit of type 1 pili	52
Ribosomal DNA-based primers		
ITSF 5' GTCGTAAACAAGGTAGCCGTA	16S-23S rRNA intergenic	53
ITSReub 5' GCCAAGGCATCCACC	transcribed spacers	54
27f 5' AGAGTTTGATCMTGGCTCAG	Conserved 16S rDNA used to	55
342r 5' CTGCTGCSYCCCGTAG	amplify ribosomal genes	56

Statistical analysis: For statistical analysis of data we applied a Chi square test based on the Mantel Haenszel method (Epi Info version 6.04, CDC, Atlanta, Georgia, USA).

RESULTS

5 A total of 84 biopsies from 15 controls, 13
Crohn's disease (CD) patients, and 19 ulcerative colitis
(UC) patients were undertaken (table 1). In most cases, more
than one biopsy was obtained from multiple sites, or
adjacent sites from the same subject. DNA was extracted from
10 each biopsy sample and subjected to RISA analysis (fig 1).
We were able to identify bands (~450 bp) that were
consistently present in approximately 70% of patients but in
less than 30% of controls (fig 1). The bands in the controls
were also of a much lower density than those from IBD
15 tissue. Five bands from IBD tissue were retrieved from the
gel, sequenced, and aligned with Genbank sequence and found
to be *E. coli*.

 Resuscitation of biopsies by resuspension in
buffered peptone water, incubation for 16 h at 37°C, and
20 subsequent plating on chromogenic agar allowed growth of
coliform bacteria in over 90% of biopsy tissues. Chromogenic
agar enriches for predominantly coliform bacteria but only
lactose-fermenting bacteria (*E. coli*) turn purple. We could
culture purple colonies in only 46.7% of control subjects,
25 69.2% of CD patients and 63.2% of UC patients even though
non-*E. coli* coliforms (white and pink colonies) could be
cultured from most biopsies (table 3).

Serial dilution and plating on chromogenic agar allowed for quantification of total coliforms and *E. coli* (fig 2). The numbers of *E. coli* ($4 \times 10^2/\text{ml}$) and non-*E. coli* coliforms ($6.3 \times 10^5/\text{ml}$) cultured were higher ($p < 0.05$) in CD and UC than in controls. There was a poor correlation ($r = 0.22$) between site of inflammation and presence of *E. coli* and data was pooled by subject.

Table 3 Microbial tissue phenotypes encountered from control and IBD tissues^a.

Disease	Patient	<i>E. coli</i> ^b group	Non- <i>E. coli</i> ^c	SPATE	<i>E. coli</i> adhesins	<i>E. coli</i> H-type	<i>E. coli</i> type 1 fimbriae
Control	41A	ND	Enterobacter sp. Bacteroides fragilis Bft1 ^d	ND	-	-	-
Control	43A	B2		Vat	Ag43, PAIH7 I		bovine
Control	48B	B2		Vat, Pic-like	Sfp ^e	H4	APEC
Control	50A	B2		Vat, Sat, Pic	Ag43	H1	APEC
Control	58	ND	Enterobacter flavescens	ND	-	-	-
Control	59B	ND	Klebsiella oxytoca	ND	-	-	-
Control	73A	A	Klebsiella sp.	ND	AIDA-I	H12	bovine
Control	69A	ND	Enterococcus sp. Bacteroides fragilis Bft1	ND	-	-	-
Control	76	ND	Escherichia fergusonii Staphylococcus epidermidis Staphylococcus capitis Bacteroides fragilis Bft2 ^d	SigA, Sat ^g , SepA-		-	-
Control	80	ND	Escherichia fergusonii Klebsiella sp.	SigA, Sat ^g , SepA-		-	-
Control	81	ND	Enterococcus durans	ND	-	-	-
Control	87	ND	Klebsiella sp.	ND	-	-	-
Control	90	A		ND	AIDA-I, Ag43	H11	bovine
Control	17B	D		ND	Ag43	H4	UPEC
Control	88	A	Klebsiella sp.	ND	AIDA-I	H21	bovine
CD	15B	B2		Vat, Pic-like	Ag43	H4	APEC
CD	79A	B2		Vat, Sat, Pic	Ag43	H1	APEC
CD	91A	B1	Acidovorax sp.	ND	AIDA-I	H21	bovine
CD	118	ND	Staphylococcus auricularis	ND	-	-	-
CD	120A	D		EspI	PAI I, AIDA-I, Ag43	H52	bovine
CD	124A	ND	Escherichia fergusonii	SigA, Sat ^g , SepA-		-	-

			Enterococcus faecium				
			Bacteroides fragilis Bft1				
CD	125B	ND	Klebsiella oxytoca	ND	-	-	-
CD	126C	B2		ND	AIDA-I, Ag43	H39	bovine
CD	132	B2		Sat	AfaE-3, Ag43, AIDA-I	H4	bovine
CD	137	ND	Klebsiella sp.	ND	-	-	-
CD	138	ND	Staphylococcus aureus	ND	-	-	-
			Pseudomonas putida				
			Klebsiella sp.				
CD	146	B2		Vat	Sfa ^f , Ag43, PAI I	H7	APEC
CD	149	B2		Vat, Pic	Sfa, Ag43, PAI I	H5	APEC
CD	128	B2	Escherichia fergusonii	Vat, SigA, Sat ^g , SepA	Sfa	H7	UPEC
UC	117A	D		ND	AIDA-I, Ag43	H18	bovine
UC	119A	ND	Enterococcus faecium	ND	-	-	-
UC	121	B1		ND	ND	H10	bovine
UC	122	ND	Enterococcus faecalis	ND	-	-	-
			Bacteroides fragilis Bft1				
UC	127A	B2	Enterobacter cloacae	Vat, Sat, Pic	F1C	H7	bovine
UC	130	ND	Bacillus licheniformis	ND	-	-	-
			Bacillus pumilus				
UC	131A	B2	Escherichia fergusonii	Vat, Sat ^g , Pic, SigA, Sat, SepA	Ag43, PAI H10		APEC
			Bacteroides fragilis Bft2				
UC	133A	ND	Klebsiella sp.	ND	-	-	-
UC	135	B2	Klebsiella oxytoca	Pic	AIDA-I, Ag43	H31	bovine
UC	136	B2	ND	Vat	Ag43, PAI I	H7	bovine
UC	139	D	ND	ND	AIDA-I	H6	APEC
UC	140A	ND	Enterococcus sp.	ND	-	-	-
UC	141	B2	ND	ND	Sfa, Ag43, PAI I	H7	APEC
UC	142	B2	ND	Vat	AfaE-3, Ag43	H5	APEC
UC	143	B2	ND	Sat, Pic-like	AfaE-3, Ag43	H4	APEC
UC	145B	B2	ND	Vat, Pic-like	Ag43	H4	bovine
UC	147	B2	Klebsiella sp.	Vat	Sfa, Ag43	H5	APEC
UC	134A	ND	Klebsiella sp.	ND	-	-	-

^aA total of 84 biopsies from 47 subjects were used. Data are not shown by site, and were pooled, because the correlation between bacterial species and site, within and between subjects was low ($r = 0.22$).

^bA total of 28 *E. coli* strains and 34 non-*E. coli* were picked and assessed. ND = *E. coli* not detected.

^cIdentity was determined by sequence the 16S rDNA between with 27f and 342r.

^dBft1, Bft2 are enterotoxins produced by *Bacteroides fragilis* detected from tissue by nested-PCR method.

^esfp is a gene cluster of Sorbitol-Fermenting Enterohemorrhagic *E. coli* O157:H⁻.

^fsfa gene is coding for S-fimbriae minor subunit of *E. coli* UTI89.

^gSat - gene coding for autotransporteur toxin Sat identified in *E. fergusonii*.

Five purple colonies were picked from all agar plates (total 135 colonies) that showed *E. coli* growth. Taxonomic identity of the *E. coli* was confirmed by

5 reactivity to indole, methyl red, Vogues Proskauer, and citrate utilisation. There were also few polymorphisms in the RISA product, confirming that isolates were predominantly clonal by subject, irrespective of site. We selected 150 colonies of non-*E. coli* (5 per biopsy) of pink,

10 blue, and white colonies from plates and were able to confirm that none of these colonies were *E. coli* by checking for reactivity to indole, methyl red, Vogues Proskauer, and citrate utilisation. The identities of these colonies were determined by 16S-rDNA sequence analysis (table 3).

15 *E. coli* comprises four phylogenetic groups (A, B1, B2, and D) with virulent types typically belonging to groups B2 and D.15 These groups can be identified by a simple PCR procedure of the chuA, and yjaA genes and a cryptic DNA fragment (table 2). The authors15 indicated that these

20 groups could be identified in a single multiplex PCR for all DNA targets but we did not get reproducible results. We modified the primers (table 2) and the PCR conditions and were able to generate highly reproducible results from a suite of pathogenic and non-pathogenic *E. coli* in our

25 laboratory (data not shown). The abundance of the pathogenic

B2 + D groups was significantly ($p = 0.04$) greater in IBD than in controls (table 4).

E. coli were also assayed for the presence of serine protease autotransporter proteins (SPATE). Alignments were made of prominent group (Vat, Sat, Pic, EspI) of SPATE nucleic acid sequences, and primers to conserved regions were designed to amplify targets which, when digested with HaeIII, were diagnostic of the different groups of SPATE. Identity of each SPATE was confirmed by sequence analysis. One new SPATE, Pic-like, was identified by sequence analysis and alignment with known SPATE. When all SPATE-positive E. coli isolates were totalled, IBD patients had a higher number of SPATE-positive isolates than the controls (table 4). Interestingly this SPATE sequence appeared in the enteropathogenic E. coli E22 genome (Genbank: AAJV01000028) but no functions were assigned.

Table 4 Distribution of virulence features of *E. coli* isolated from IBD patients and controls

Item ¹	Control (15 subjects 28 tissues)	UC (19 patients 27 tissues)	CD (13 patients 29 tissues)
Group			
A	3 (20) ²	0	0
B1	0	1 (5.3)	1 (7.7)
B2	3 (20)	9 (47.4)	7 (53.8)
D	1 (6.7)	2 (10.5)	1 (7.7)
B2+D	4 (26.7)	11 (57.9)	8 (61.5)
SPATE			
Vat	3 (20)	7 (36.8)	4 (30.8)
Pic	1 (6.7)	3 (15.8)	3 (23.1)
Sat	3 (20)	4 (21.05)	3 (23.1)
Pic-like	1 (6.7)	2 (15.4)	1 (7.7)
EspI	0	0	1 (7.7)
SepA	2 (13.3)	2 (10.5)	1 (7.7)
SigA	2 (13.3)	2 (10.5)	1 (7.7)
Total	5 (33.3)	9 (47.4)	7 (53.8)
Adhesins			
Ag43 ³	4 (26.7)	9 (47.4)	7 (53.8)
AfaE3	0	2 (10.5)	1 (7.7)
AIDA-I	3 (20)	3 (15.8)	4 (30.8)
F1C	0	1 (5.3)	0
Sfa	0	3 (15.8)	2 (15.4)
PAI I	1 (6.7)	3 (15.8)	3 (23.1)
Sfp	1 (6.7)	0 (5.3)	0
Total	7 (46.6)	12 (63.2)	8 (61.5)

¹Functions of each of the genotypic characteristics is given in table 2.

²Values in parentheses represent the frequency of items in patients as percentage.

³Antigen 43 (Ag43) is a self-recognizing surface adhesin found in most *Escherichia coli* strains.

The PCR analyses of *E. coli* isolates for a range of adhesins commonly found in pathogenic *E. coli* were conducted (tables 3, 4). Isolates were only positive for agn43, aidA, and gene clusters coding for AfaE-3, F1C, Sfa, PAI I, and Sfp. No *E. coli* positive for cnf1, cnf2, eae, hlyA, ipgB, aggR, bmaE, or genes coding for verotoxins, heat-stable and heat-labile toxins were found in biopsies of patients with UC or CD.

Chi-squares test based on the Mantel-Haenszel method (table 5) was performed to verify relationships between disease, B2 + D genotype, SPATE toxins, and adhesins. We observed a significant difference ($p = 0.04$) in the number of *E. coli* isolates from the B2 + D genotype for both UC and CD patients (Table 5). The relationship between IBD and an adhesin ($p = 0.07$), or two or more adhesins ($p = 0.03$) was significant but not with a SPATE plus an adhesin ($p = 0.15$). The relationship between the B2 + D genotype and SPATE9(s), adhesion(s), or APEC was significant ($p < 0.05$).

Table 5 Statistical relationship between the presence of virulence factors in IBD patients and control subjects

Disease	Virulence factor	p-value
IBD	2 or more adhesins	0.03*
IBD	B2 + D	0.04*
IBD	Adhesins	0.07
IBD	SPATE(s) + adhesin(s)	0.15
B2 + D genotype	Virulence feature	p-value
B2 + D	SPATE(s) + adhesin(s)	0.002**
B2 + D	Adhesins	0.04*
B2 + D	SPATE	0.02*
B2 + D	APEC	0.04*

Statistically significant results *($p < 0.05$), **($p < 0.01$)

DISCUSSION

In our laboratory work, we: (a) use a culture-independent method to identify DNA bands that are present in IBD tissue but not in controls; (b) cut the bands out of the gel and sequence them; (c) identify the bacterial species based on the sequence composition; (d) specifically culture the bacterial group identified from the sequence information, and (e) investigate potential virulence factors

in the cultured bacteria. The RISA analysis (fig 1) provided the rationale for culturing *E. coli* and other Enterobacteriaceae.

The numbers of Enterobacteriaceae in our biopsy tissues were 3 to 4 logs higher in the IBD tissue than in the controls (fig 2). Martin et al (ref 17) only found a significant increase in the numbers of *E. coli* in CD biopsies, but not UC, after a mucin-releasing step with dithiothreitol. Darfeuille-Michaud et al (ref 20) isolated *E. coli* in higher numbers from IBD tissue than from controls but did not enumerate the bacteria on the epithelial tissue with techniques to determine numbers in a range greater than one log. Mylonaki et al (ref 21) used fluorescence microscopy to demonstrate that the numbers of *E. coli* were high in rectal tissue of UC patients but not of controls, but this methodology did not allow for decimal enumeration.

Conte et al (ref 22) could demonstrate that gram-negative bacteria, including *E. coli* increased by 3 to 4 logs in IBD tissue, a result strikingly similar to ours. It is thus clear that the numbers of Enterobacteriaceae do increase on the epithelial tissue of IBD patients, irrespective of the differences in techniques used. The major difference in our study was that we resuscitated tissue in buffered peptone water to ensure that bacteria, even at very low numbers, were given the maximum chance to grow. This was done given the fact that *E. coli* in environmental samples enter into the viable, but non-culturable state, making them difficult to grow without a resuscitation step (ref 23).

Clermont et al (ref 15) developed a method to type pathogenic *E. coli* using *chuA*, a gene required for heme

transport in enterohaemorrhagic O157:H7 *E. coli*, *yjaA*, a gene identified in *E. coli* K-12 but which has no known function, and TSPE4.C2, a cryptic fragment that was identified from subtractive libraries. These genes, when applied to 230 isolates, determined that types B2, and to a lesser extent D, included virulent extraintestinal strains of *E. coli* but the prevalence of B2 and D in gastrointestinal isolates was not determined. *E. coli* from stool samples from a range of geographically separate normal healthy human subjects determined that non-pathogenic groups A and B1 were most prevalent, while group D only made up 15%, and group B2 11% of isolates.²⁴ A significant relationship ($p < 0.05$) between the B2 genotype of "resident", or adherence factor carrying *E. coli* in infants²⁵ and in human colonic cells of adults²⁶ has been demonstrated. We demonstrated a significant relationship ($p = 0.04$) between IBD and the B2 + D genotype (table 5). Our data suggest a significant relationship between the presence of a SPATE ($p = 0.02$), or an adhesin ($p = 0.04$), in a B2 + D positive *E. coli* strains.

The "resident" population that Nowrouzian et al (ref 25) referred to are strains that have adherence factors, in particular P fimbriae that promote adherence to enterocytes. This definition is of course somewhat arbitrary because there are a large number of cell factors that promote adhesin to intestinal tissue. *E. coli* from our controls had relatively few adhesins (table 4), while IBD strains had a higher prevalence of Ag43, AIDA-I, Sfa, AfaE-3, and PAI I. Ag43 is a surface adhesin that promotes bacterial biofilm formation due to cell-to-cell aggregation (ref 27), Sfa is one of a class of S-fimbrial adhesins (ref 16), and AIDA-I is an adhesin-like protein (ref 16). Martin et al (ref 17) measured adherence and invasion of IBD-

derived *E. coli* in tissue culture, and showed that *E. coli* strains isolated from Crohn's disease patients possessed haemagglutinating ability to all red cells regardless of blood group. We did not measure adherence and invasion in a cell culture assay but all *E. coli* isolates both from IBD tissues and control group were negative for the presence of *bmaE* gene encoding M-agglutinin.

A novel finding of this study is the higher prevalence of *Escherichia coli* from the B2+D phylogenetic group in IBD tissues. Moreover, SPATE (ref 28) plus adhesins which have been primarily associated with *E. coli* isolated from urinary tract infections, were also more prevalent in *E. coli* isolates from IBD patients. SPATE is a unique class of transporter found in the Enterobacteriaceae that direct their own transport across the outer cell membrane. These proteins have been implicated in virulence but their precise role is not known. We believe that the functional properties of SPATE make them potentially important in IBD because they exhibit functions like degradation of the barrier function of the gut, and cleavage of proteins in the enterocyte, all of which are phenotypes associated with IBD (ref 28). For example Vat, Pic, and Pic-like have haemagglutinin, mucinase, and elastase activity (ref 29, 30), Sat has cytotoxic effects (ref 31-33) on cells as well as elastase activity (ref 27-30), and EspP cleaves lipoproteins (ref 27-30). More recently, functional studies on SPATE have demonstrated that Pic and Sat can cleave coagulation factor V, potentially linking it to haemorrhagic events in the gut. Additionally, Pic is thought be involved in colonisation of *E. coli* to intestinal tissue (ref 34).

One means by which SPATE may be involved in promoting inflammation could be as an accessory protein in

pathogenicity islands. EspC, a SPATE from enteroaggregative *E. coli* (EAEC) (ref 35) and the non-SPATE Tsh autotransporter (ref 36) have been associated with PAI. Bidet et al (ref 37) suggested that PAI I_{J96} were associated with hra, hlyA, cnf1, and pap. We determined the prevalence of hlyA and cnf1 and cnf2 in our isolates but they were not present. When we designed primers that were conserved for a range of PAI, only 23.1% of CD and 15.8% of UC were positive. If these were the only PAI, then less than 50% of our SPATE would be associated with PAI. Future work would investigate the possibility of the association of other PAI with SPATE.

One might argue that IBD is caused by a dysfunctional immune system that results in an increase in *E. coli* on the gut tissue, and that *E. coli* has nothing whatever to do with initiating inflammation in UC or CD. If this was true then the genotypes of *E. coli* isolated from tissue should be a consequence of random colonisation of gut mucosa. Consequently the distribution of these genotypes should be equivalent among IBD and control tissues. This is not the case, with IBD having more ($p = 0.04$) putatively pathogenic bacteria than controls. What we do not know at this point is whether the increase in certain types of *E. coli* in IBD is a consequence of inflammation, or the cause. One way forward is to describe the *E. coli* population before inflammation sets in, for example, obtain biopsy tissue that ranges along the colon, distal and proximal to UC lesions.

The present invention is not to be limited in scope by the specific embodiments described herein. Various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying

figures. Such modifications are intended to fall within the scope of the claims.

The present invention is directed to each individual feature, system, material and/or method described herein. In addition, any combination of two or more such features, systems, materials and/or methods, provided that such features, systems, materials and/or methods are not mutually inconsistent, is included within the scope of the present invention.

10 In the specification and claims, all transitional phrases or phrases of inclusion, such as "comprising," "including," "carrying," "having," "containing," "composed of," "made of," "formed of," "involving" and the like shall be interpreted to be open-ended, i.e. to mean "including but
15 not limited to" and, therefore, encompassing the items listed thereafter and equivalents thereof as well as additional items. Only the transitional phrases or phrases of inclusion "consisting of" and "consisting essentially of" are to be interpreted as closed or semi-closed phrases,
20 respectively. The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one." The expression "A or B", unless clearly indicated to the contrary, should be understood to
25 mean "A or B or both".

Various publications are cited herein, the disclosures of which are incorporated by reference in their entirety or in pertinent part, as is understood from the context of the publication being cited. In cases where the
30 present specification and a document incorporated by reference and/or referred to herein include conflicting

disclosure, and/or inconsistent use of terminology, and/or the incorporated/referenced documents use or define terms differently than they are used or defined in the present specification, the present specification shall control.

5 REFERENCES

1. Koutroubakis IE. Therapy insight: vascular complications in patients with inflammatory bowel disease. *Nat Clin Pract Cardiovasc Med* 2005;2:266-272.
2. Sartor RB. Role of commensal enteric bacteria in the
10 pathogenesis of immune-mediated intestinal inflammation:
lessons from animal models and implications for
translational research. *J Pediatr Gastroenterol Nutr* 2005;40
(Suppl 1):S30-S31.
3. Rath HC. The role of endogenous bacterial flora:
15 bystander or the necessary prerequisite? *Eur J Gastroenterol
Hepatol* 2003;15:615-620.
4. Bernstein CN, Wajda A, Blanchard JF. The clustering of
other chronic inflammatory diseases in inflammatory bowel
disease: a population-based study. *Gastroenterol*
20 2005;129:827-836.
5. Garcia-Rodriguez LA, Gonzalez-Perez A, Johansson S, et
al. Risk factors for inflammatory bowel disease in the
general population. *Aliment Pharmacol Ther* 2005;22:309-315.
6. Hume G, Radford-Smith, GL. The pathogenesis of Crohn's
25 disease in the 21st century. *Pathology* 2002;34:561-567.
7. Silverberg MS, Satsangi J, Ahmad T, et al. Toward an
integrated clinical, molecular and serological
classification of inflammatory bowel disease: Report of a

Working Party of the 2005 Montreal World Congress of Gastroenterology. Can J Gastroenterol 2005; Suppl A:5-36.

8. Korzenik JR. Past and current theories of etiology of IBD: toothpaste, worms, and refrigerators. J Clin Gastroenterol Nutr 2005;40 (Suppl 1):S30-S31.

9. Bull TJ, McMin E, Sidi-Boumedine K, Skull A, et al. Detection and verification of Mycobacterium avium subsp. paratuberculosis in fresh ileocolonic mucosal biopsy specimens from individuals with and without Crohn's disease. J Clin Microbiol 2003;41:2915-2923.

10. Eckburg PB, Bik EM, Bernstein CN, et al. Diversity of the human intestinal microbial flora. Science 2005; 308:1635-1638.

11. Riesenfeld CS, Schloss PD, Handelsman J, et al. Metagenomics: genomic analysis of microbial communities. Annu Rev Genet 2004;38:525-552.

12. Rappe MS, Giovannoni SJ. The uncultured microbial majority. Annu Rev Microbiol 2004;57:369-394.

13. Cardinale M, Brusetti L, Quatrini P, et al. Comparison of different primer sets for use in automated ribosomal intergenic spacer analysis of complex bacterial communities. Appl Environ Microbiol. 2004;70:6147-56.

14. Bernstein CN, Nayar G, Hamel A, et al. Study of animal-borne infections in the mucosas of patients with inflammatory bowel disease and population-based controls. J Clin Microbiol 2003;41:4986-4990.

15. Clermont O, Bonacorsi S, Bingen E, et al. Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl Environ Microbiol* 2000;66:4555-4558.
16. Schmidt H, Hensel M. Pathogenicity islands in bacterial pathogenesis. *Clin Microbiol Rev* 2004;17:14-56.
17. Martin HM, Campbell BJ, Hart CA, et al. Enhanced *Escherichia coli* adherence and invasion in Crohn's disease and colon cancer. *Gastroenterology*. 2004;127:80-93.
18. Machado J, Grimont F, Grimont PA, et al. Identification of *Escherichia coli* flagellar types by restriction of the amplified *fliC* gene. *Res Microbiol*. 2000;151:535-546.
19. Lane DJ. 16S/23S rRNA sequencing. *Nucleic acid techniques in bacterial systematics*. E. Stackebrandt and M. Goodfellow, eds. New York, NY, John Wiley and Sons: 115-175.
20. Darfeuille-Michaud A, Boudeau J, Bulois P, et al. High prevalence of adherent-invasive *Escherichia coli* associated with ileal mucosa in Crohn's disease. *Gastroenterology* 2004;127:412-421.
21. Mylonaki M, Rayment NB, Rampton DS, et al. Molecular characterization of rectal mucosa-associated bacterial flora in inflammatory bowel disease. *Inflamm Bowel Dis* 2005;11:481-487.
22. Conte MP, Schippa S, Zamboni I, et al. Gut-associated bacterial microbiota in paediatric patients with inflammatory bowel disease. *Gut* 2006 Apr 28; [Epub ahead of print].

23. Nystrom T. Not quite dead enough: on bacterial life, culturability, senescence, and death. Arch Microbiol 2001;176:159-164.
24. Duriez P, Clermont O, Bonacorsi S, et al. Commensal
5 Escherichia coli isolates are phylogenetically distributed among geographically distinct human populations. Microbiology 2001;147:1671-1676.
25. Nowrouzian FL, Wold AE, Adlerberth I. Escherichia coli strains belonging to phylogenetic group B2 have superior
10 capacity to persist in the intestinal microflora of infants. J Infect Dis 2005;191:1078-1083.
26. Nowrouzian FL, Adlerberth I, Wold AE. Enhanced persistence in the colonic microbiota of Escherichia coli strains belonging to phylogenetic group B2: role of
15 virulence factors and adherence to colonic cells. Microbes Infect 2006;8:834-840.
27. Henderson IR, Navarro-Garcia F, Desvaux M, et al. Type V Protein Secretion Pathway: the Autotransporter Story. Microbiol Mol Biol Rev 2004;68:692-744.
- 20 28. Henderson IR, Nataro JP. Virulence functions of autotransporter proteins. Infect Immun 2001;69:1231-1243.
29. Parham NJ, Pollard SJ, Desvaux M, et al. Distribution of the serine protease autotransporters of the Enterobacteriaceae among extraintestinal clinical Isolates
25 of Escherichia coli. J Clin Microbiol 2005;43:4076-4082.
30. Dutta PR, Cappello R, Navarro-Garcia F, et al. Functional comparison of serine protease autotransporters of enterobacteriaceae. Infect Immun 2002;70:7105-7113.

31. Betis F, Brest P, Hofman V, et al. The Afa/Dr adhesins of diffusely adhering *Escherichia coli* stimulate interleukin-8 secretion, activate mitogen-activated protein kinases, and promote polymorphonuclear transepithelial migration in T84 polarized epithelial cells. *Infect Immun.* 2003;71:1068-74.
32. Taddei CR, Fasano A, Ferreira AJ, et al. Secreted autotransporter toxin produced by a diffusely adhering *Escherichia coli* strain causes intestinal damage in animal model assays. *FEMS Microbiol Lett.* 2005; 250:263-9.
33. Guignot J, Chaplais C, Coconnier-Polter MH, et al. The secreted autotransporter toxin, Sat, functions as a virulence factor in Afa/Dr diffusely adhering *Escherichia coli* by promoting lesions in tight junction of polarized epithelial cells. *Cell Microbiol.* 2006; [Epub ahead of print]
34. Henderson IR, Czeczulin J, Eslava C, et al. Characterization of pic, a secreted protease of *Shigella flexneri* and enteroaggregative *Escherichia coli*. *Infect Immun* 1999;67:5587-5596.
35. Mellies JL, Navarro-Garcia F, Okeke I, et al. EspC pathogenicity island of enteropathogenic *Escherichia coli* encodes an enterotoxin. *Infect Immun* 2001;69: 315-324.
36. Dozois CM, Dho-Moulin M, Bree A, et al. Relationship between the Tsh autotransporter and pathogenicity of avian *Escherichia coli* and localization and analysis of the Tsh genetic region. *Infect Immun* 2000;68:4145-4154.
37. Bidet P, Bonacorsi S, Clermont O, et al. Multiple insertional events, restricted by the genetic background, have led to acquisition of pathogenicity island IIJ96-like

domains among *Escherichia coli* strains of different clinical origins. *Infect Immun* 2005;73:4081-4087.

38. Maroncle et al. *Infect Imm* 2006; 74(11): 6124-6134.

39. Sherlock et al. *J Bacteriology* 2006; 188(5):1798-1807.

CLAIMS:

1. A method for diagnosing inflammatory bowel disease (IBD) or determining susceptibility to developing IBD in a subject, the method comprising the step of assaying for
5 serine protease autotransporter (SPATE) or antigen 43 (Ag43) or both, in an enteric bacteria-containing sample from the subject, wherein the presence of SPATE in the sample indicates that the subject has IBD or is susceptible to developing IBD.
- 10 2. The method according to claim 1 wherein the subject has symptoms of IBD, or is suspected of having IBD.
3. The method according to claim 1 wherein the IBD is ulcerative colitis (UC) or Crohn's disease (CD).
4. The method according to any one of claims 1 to 3,
15 wherein SPATE is assayed by detecting SPATE nucleic acid and wherein Ag43 is assayed by detecting Ag43 nucleic acid.
5. The method according to claim 4 wherein the assay for SPATE nucleic acid or for Ag43 nucleic acid comprises detecting a region of SPATE or a region of Ag43 which is
20 conserved among enteric bacteria.
6. The method according to claim 4 wherein the assay for SPATE nucleic acid or for Ag43 nucleic acid comprises detecting a region of SPATE or a region of Ag43 which is conserved among enteric E. coli.
- 25 7. The method according to claim 4 wherein the assay for SPATE nucleic acid or for Ag43 nucleic acid comprises detecting a region of SPATE or a region of Ag43 which is specific to enteric E. coli.

8. The method according to claim 4 wherein the assay for SPATE nucleic acid or for Ag43 nucleic acid comprises detecting a region of SPATE or a region of Ag43 which is conserved among virulent enteric strains of E. coli.

5 9. The method according to claim 4 wherein the assay for SPATE nucleic acid or for Ag43 nucleic acid comprises detecting a region of SPATE or a region of Ag43 which is specific to virulent enteric strains of E. coli.

10. The method according to claim 4 wherein the assay
10 for SPATE nucleic acid or for Ag43 nucleic acid comprises detecting a region of SPATE or a region of Ag43 which is specific to E. coli of the B2 or D genotype or the B2+D genotype.

11. The method according to any one of claims 4 to 10
15 comprising detecting SPATE nucleic acid or detecting Ag43 nucleic acid, or both, by polymerase chain reaction (PCR).

12. The method according to claim 1 comprising assaying for enteric bacterial SPATE or enteric bacterial Ag43 polypeptide or both.

20 13. The method according to claim 12 wherein the SPATE or Ag43 or both are of enteric E. coli.

14. The method according to claim 13 wherein the SPATE or Ag43 or both are of virulent enteric strains of E. coli.

15. The method according to claim 13 wherein the SPATE
25 or Ag43 or both are of E. coli of the B2 or D genotype or the B2+D genotype.

16. The method according to claim 15 comprising an immunoassay for the SPATE or the Ag43 polypeptide or both.

17. The method according to any one of claims 1 to 16 comprising the steps of:

(i) contacting an enteric bacteria-containing sample from a subject with an antibody immunospecific
5 against the SPATE or against the Ag43, or with antibodies against both SPATE and Ag43, under conditions suitable to form a complex between the SPATE or the Ag43 and the antibody; and

(ii) detecting presence or absence of the complex;
10 wherein presence of the complex indicates that the subject has or is susceptible to developing IBD.

18. The method according to any one of claims 1 to 17 wherein the level of SPATE or Ag43 or both in the sample is determined, whereby a higher level of SPATE or Ag43 or both
15 compared to a cut-off value indicates that the subject has IBD or or is susceptible to developing IBD.

19. The method according to any one of claims 1 to 18 wherein the subject is human.

20. The method according to any one of claims 1 to 19
20 wherein the sample is a gut tissue biopsy, a sample of intestinal mucosa, stool sample, or intestinal wash.

21. The method according to claim 20 wherein the sample is a colonoscopy tissue biopsy from the lower gastrointestinal (GI) tract.

25 22. A method for evaluating effectiveness of a treatment for inflammatory bowel disease (IBD) in a subject, the method comprising the steps of:

(i) determining serine protease autotransporter (SPATE) level or Ag43 level, or both, in an enteric bacteria-containing sample from a treated subject having IBD, and

5 (ii) comparing the SPATE level or the Ag43 level or both from step (i) to SPATE and Ag43 levels determined in an enteric bacteria-containing sample from an untreated subject having IBD,

wherein the untreated and treated subjects are either
10 different subjects or are the same subjects before and after undergoing the treatment, and wherein a lower level of SPATE or Ag43 or both in treated subjects indicates that the treatment is effective for IBD.

23. The method according to claim 22 wherein the IBD
15 is ulcerative colitis (UC) or Crohn's disease (CD).

24. The method according to claim 22 or 23 wherein the SPATE level is determined by assaying for SPATE nucleic acid or Ag43 nucleic acid or both.

25. The method according to claim 24 wherein the assay
20 for SPATE or Ag43 nucleic acid comprises detecting a region of SPATE or a region of Ag43 which is conserved among enteric bacteria.

26. The method according to claim 24 wherein the assay
for SPATE or Ag43 nucleic acid comprises detecting a region
25 of SPATE or a region of Ag43 which is conserved among enteric E. coli.

27. The method according to claim 24 wherein the assay
for SPATE or Ag43 nucleic acid comprises detecting a region

of SPATE or a region of Ag43 which is specific to enteric E. coli.

28. The method according to claim 24 wherein the assay for SPATE or Ag43 nucleic acid comprises detecting a region
5 of SPATE or a region of Ag43 which is conserved among virulent enteric strains of E. coli.

29. The method according to claim 24 wherein the assay for SPATE or Ag43 nucleic acid comprises detecting a region of SPATE or a region of Ag43 which is specific to virulent
10 enteric strains of E. coli.

30. The method according to claim 24 wherein the assay for SPATE or Ag43 nucleic acid comprises detecting a region of SPATE or a region of Ag43 which is specific to E. coli of the B2+D genotype.

15 31. The method according to any one of claims 24 to 30 comprising detecting SPATE nucleic acid or detecting Ag43 nucleic acid or both by polymerase chain reaction (PCR).

32. The method according to claim 22 wherein the level of enteric bacterial SPATE or enteric bacterial Ag43 is
20 determined by assaying for SPATE polypeptide or Ag43 polypeptide or both.

33. The method according to claim 32 wherein the SPATE or Ag43 or both are of enteric E. coli.

34. The method according to claim 33 wherein the SPATE
25 or Ag43 or both are of virulent enteric strains of E. coli.

35. The method according to claim 33 wherein the SPATE or Ag43 or both are of E. coli of the B2 or D genotype or the B2+D genotype.

36. The method according to claim 35 comprising an immunoassay for the SPATE or the Ag43 polypeptide or both.

37. The method according to any one of claims 22 to 36 wherein the subject is human.

5 38. The method according to any one of claims 22 to 37 wherein the sample is a gut tissue biopsy, a sample of intestinal mucosa, stool sample, or intestinal wash.

39. The method according to claim 38 wherein the sample is a colonoscopy tissue biopsy from the lower
10 gastrointestinal (GI) tract.

40. A commercial package for diagnosing inflammatory bowel disease (IBD) or determining susceptibility to developing IBD in a subject, the package comprising an agent for detecting serine protease autotransporter (SPATE) or for
15 detecting Ag43, or both, in an enteric bacteria-containing sample from the subject, and instructions for using the agent to detect SPATE or Ag43 or both in the sample, thereby diagnosing or prognosing IBD in the subject.

41. The commercial package according to claim 40
20 wherein the subject has symptoms of IBD, or is suspected of having IBD.

42. The commercial package according to claim 40 wherein the IBD is ulcerative colitis (UC) or Crohn's disease (CD).

25 43. The commercial package according to any one of claims 40 to 42, wherein the agent is for detecting SPATE nucleic acid or Ag43 nucleic acid.

44. The commercial package according to claim 43 wherein the agent detects a region of SPATE or a region of Ag43 which is conserved among enteric bacteria.

45. The commercial package according to claim 43
5 wherein the agent detects a region of SPATE or a region of Ag43 which is conserved among enteric E. coli.

46. The commercial package according to claim 43 wherein the agent detects a region of SPATE or a region of Ag43 which is specific to enteric E. coli.

10 47. The commercial package according to claim 43 wherein the agent detects a region of SPATE or a region of Ag43 which is conserved among virulent enteric strains of E. coli.

48. The commercial package according to claim 43
15 wherein the agent detects a region of SPATE or a region of Ag43.

49. The commercial package according to claim 43 wherein the agent detects a region of SPATE or a region of Ag43 which is specific to E. coli of the B2 or D genotype or
20 the B2+D genotype.

50. The commercial package according to any one of claims 43 to 49 wherein the agent detects SPATE nucleic acid or Ag43 nucleic acid by polymerase chain reaction (PCR).

51. The commercial package according to any one of
25 claims 40 to 42 wherein the agent is for detecting SPATE polypeptide or Ag43 polypeptide or both.

52. The commercial package according to claim 51 wherein the agent is for detecting SPATE polypeptide or Ag43 polypeptide in an immunoassay.

53. The commercial package according to claim 52
5 wherein the agent is an antibody which is immunospecific against SPATE or against Ag43 of enteric bacteria.

54. The commercial package according to claim 52 wherein the agent is an antibody which is immunospecific against SPATE or against Ag43 of enteric E. coli.

10 55. The commercial package according to claim 52 wherein the agent is an antibody which is immunospecific against SPATE or against Ag43 of virulent enteric strains of E. coli.

56. The commercial package according to claim 52
15 wherein the agent is an antibody which is immunospecific against SPATE or against Ag43 of E. coli of the B2 or D genotype or the B2+D genotype.

57. The commercial package according to any one of claims 51 to 56 further comprising SPATE polypeptide or Ag43
20 polypeptide or both for generating standard curves.

58. A method of testing potential therapeutic agents for treating inflammatory bowel disease (IBD), the method comprising contacting a test compound with serine protease autotransporter (SPATE) polypeptide or with Ag43 polypeptide
25 and determining whether the test compound binds to SPATE or inhibits SPATE protease activity, or whether the test compound binds Ag43 or inhibits Ag43 aggregation activity or cell adhesion activity, wherein binding to SPATE or inhibition of SPATE protease activity, or binding to Ag43 or
30 inhibition of Ag43 aggregation or cell adhesion indicates

that the compound is a potential therapeutic agent for treating IBD.

59. The method according to claim 58 wherein the IBD is ulcerative colitis (UC) or Crohn's disease (CD).

5 60. The method according to claim 58 or 59 wherein the SPATE or Ag43 or both are of enteric E. coli.

61. The method according to claim 58 or 59 wherein the SPATE or Ag43 or both are of virulent enteric strains of E. coli.

10 62. The method according to claim 58 or 59 wherein the SPATE or Ag43 or both are of E. coli of the B2 or D genotype or the B2+D genotype.

63. A method for cultivating enteric bacteria, the method comprising the steps of:

15 a) resuscitating an enteric bacteria-containing sample from a subject by resuspension in buffered peptone water; and

b) culturing the resuspension.

64. The method according to claim 63 wherein the
20 enteric bacteria are E. coli.

65. The method according to claim 63 or 64 wherein the enteric bacteria-containing sample is a colonoscopy tissue biopsy, and wherein the bacteria in the sample is resuscitated without washing the tissue or removing mucus.

25 66. The method according to any one of claim 63 to 65 wherein the enteric bacteria are from subjects having IBD.

67. A method for phylogenetic identification of enteric bacteria associated with inflammatory bowel disease (IBD), the method comprising the steps of:

(i) amplifying DNA from an enteric bacteria-
5 containing sample from a subject having IBD;

(ii) amplifying DNA from an enteric bacteria-
containing sample from a subject free of IBD;

(iii) comparing the amplified DNA obtained from
step (i) with amplified DNA obtained from step (ii) to mark
10 out DNA that is associated with the subject having IBD; and

(iv) performing phylogenetic analysis to identify
bacteria containing the DNA associated with the subject
having IBD.

68. The method according to claim 67 wherein the
15 enteric bacteria-containing sample is a colonoscopy tissue
biopsy.

69. The method according to claim 67 or 68 wherein the
DNA amplification comprises PCR amplification of a
polymorphic region flanked by conserved sequences of enteric
20 bacteria, wherein the PCR amplification employs primers
specific for the conserved sequences.

70. The method according to claim 69 wherein the
primers are specific to at least a portion of the 16S rRNA
gene, or to at least a portion of the 23S rRNA gene, or to
25 at least a portion of both the 16S and the 23S rRNA genes.

71. A method for phylogenetic identification of
enteric bacteria associated with inflammatory bowel disease
(IBD), the method comprising the steps of:

(i) amplifying DNA from enteric bacteria wherein the bacteria are from a subject having IBD and were cultivated according to the method of any one of claim 5 to 5-2;

5 (ii) amplifying DNA from enteric bacteria wherein the bacteria are from an IBD-free subject and were cultivated according to the method of any one of claim 5 to 5-2;

10 (iii) comparing the amplified DNA obtained from step (i) with amplified DNA obtained from step (ii) to mark out DNA from bacteria associated with the subject having IBD; and

(iv) performing phylogenetic analysis to identify the bacteria associated with the subject having IBD.

AMENDED CLAIMS

received by the International Bureau on 04 October 2007 (04.10.2007)

(i) amplifying DNA from enteric bacteria wherein the bacteria are from a subject having IBD and were cultivated according to the method of any one of claim 63 to 65;

5 (ii) amplifying DNA from enteric bacteria wherein the bacteria are from an IBD-free subject and were cultivated according to the method of any one of claim 63 to 65;

10 (iii) comparing the amplified DNA obtained from step (i) with amplified DNA obtained from step (ii) to mark out DNA from bacteria associated with the subject having IBD; and

(iv) performing phylogenetic analysis to identify the bacteria associated with the subject having IBD.

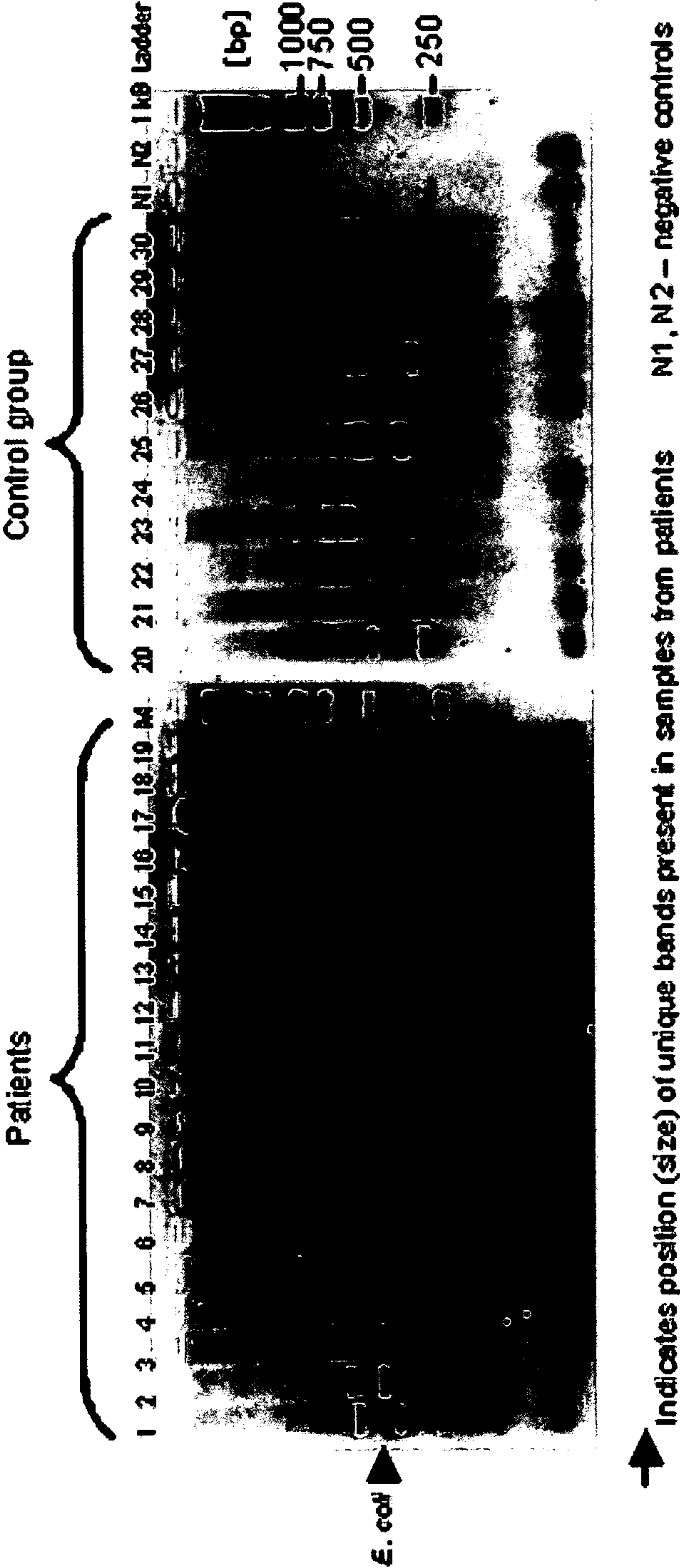


Figure 1

2 / 13

Figure 2.

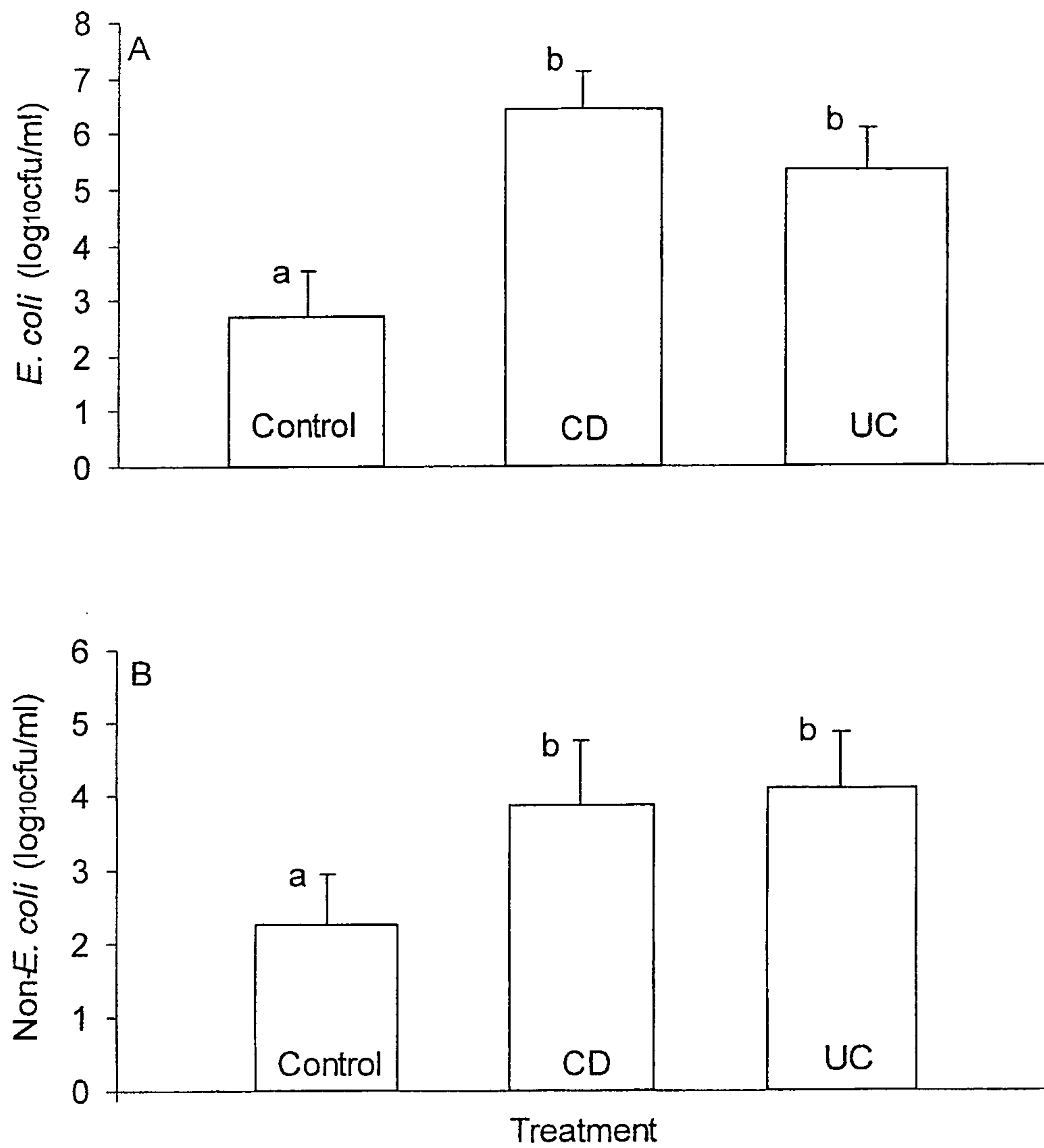


Figure 3A. Alignment of SPATE nucleotide sequences amplified by RISA with primers SPATE1 and SPATE2.

	*	20	*	40	*	60	*	80
EspP_X97542	:	GAGGTCAACAACCTGAACAAACGTA	TGGGTGACCTGCGTGATATCAACGGCGAAGCCGGTGCA	TGGGCACGCATCATGAG				
Pssa_Y13614	:	GAGGTCAACAACCTGAACAAACGTA	TGGGTGACCTGCGTGATATCAACGGCGAAGCCGGTGCA	TGGGCACGCATCATGAG				
EpeA_AY258503	:	GAAGTCAACAACCTGAACAAACGTA	TGGGTGACCTGCGTGACATCAATGGCGAAGCCGGTGCA	TGGGCACGCATCATGAG				
AJ278144	:	GAGGTCAACAACCTGAATAAACGTA	TGGGTGACCTGCGTGACATCAATGGTGAAGCTGGCGCATGGGCACGTATCATGAG					
Sat_AJ586888	:	GAGGTCAACAACCTTAACAAACGTA	TGGGTGATCTGCGTGACATTAACGGTGAGGCCGGTGCA	TGGGCCCGTATCATGAG				
Pet_AF056581	:	GAGGTCAACAACCTCAACAAACGTA	TGGGTGATCTGCGTGACATTAACGGTGAGTCCGGTGCA	TGGGCCCGGAATCATTAG				
Sat_AX702523	:	GAGGTCAACAACCTTAACAAACGTA	TGGGTGATCTGCGTGACATTAACGGTGAGTCCGGTGCA	TGGGCCCGCATCATGAG				
SigA_CP000038	:	GAAGTCAACAACCTGAATAAACGAA	TGGGGATCTGCGTGACACCAATGGCGGATGCCGGTGCT	TGGGCACGCATAATGAA				
SepA_AF348706	:	GAGGTCAATAAACCTGAACAAACG	TATGGGTGACCTGCGGGATACTCAGGGGATGCCGGTGCT	TGGGCACGCATAATGAA				
SepA_Z48219	:	GAGGTCAATAAACCTGAACAAACG	TATGGGTGACCTGCGGGATACTCAGGGGATGCCGGTGCT	TGGGCACGCATAATGAA				
EspC_AF297061	:	GAGGTCAATAAACCTGAACAAAGC	TATGGGTGACCTGCGGGATACTCAGGGGATGCCGGTGCT	TGGGCCCGCATCATGAA				
EspC_U69128	:	GAGGTCAATAATCTGAACAAAGC	TATGGGTGACCTGCGGGATACTCAGGGGATGCCGGTGCT	TGGGCCCGCATCATGAA				
SepA1_AY604009	:	GAGGTCAACAACCTGAACAAAGC	TATGGGTGACCTGCGTGATACGACGGGGATGCTG	GAGGGTGCGGTATCATGAA				
EatA_AY163491	:	GAAGTCAATAATCTGAACAAACG	TATGGGAGATTTACGGGATAGTCAGGGAGATGCTG	GAGGGTGCGGTATCATGAA				
Vat_DD002707	:	GAAGTTAACAACCTGAACAAACG	CATGGGCCGATTTAGGGATATTAATGGCGAAGCCGGT	ACGTGGGTGCGTCTGCTGAA				
Hbp_AF218073	:	GAAGTTAACAACCTGAACAAAGC	TATGGGGATCTGCGTGATACTCAGGGCGAGGACGGA	ATGTGGGTACGTATCATGAA				
Boa_AY876285	:	GAAGTCAACAACCTGAACAAACG	TATGGGTGAACCTGCGTGATACGAACCGTG	CCCTGGGCCGTATCATGAA				
Citr_rodendium	:	GAAGTTAACAACCTGAACAAACG	TATGGGTGACCTGCGTGACACAAACCGTG	CCCTGGGCCGTATCATGAG				
PicS_AE005674	:	GAAGTTAACAATCTGAACAAACG	TATGGGTGACCTGCGTGACACAAACCGTG	CCCTGGGCCGTATCATGAG				
PicEc_AX276281	:	GAAGTCAACAATCTGAACAAACG	TATGGGTGACCTGCGTGACACAAACCGTG	CCCGGAGCCTGGGCCGTATCATGAG				
EspP_X97542	:	GAGGTCAACAACCTGAACAAACG	TATGGGTGACCTGCGTGATATCAACGGCGAAGCCGGTGCA	TGGGCACGCATCATGAG				
Pssa_Y13614	:	GAGGTCAACAACCTGAACAAACG	TATGGGTGACCTGCGTGATATCAACGGCGAAGCCGGTGCA	TGGGCACGCATCATGAG				
EpeA_AY258503	:	GAAGTCAACAACCTGAACAAACG	TATGGGTGACCTGCGTGACATCAATGGCGAAGCCGGTGCA	TGGGCACGCATCATGAG				
AJ278144	:	GAGGTCAACAACCTGAATAAACG	TATGGGTGACCTGCGTGACATCAATGGTGAAGCTGGCGCATGGGCACGTATCATGAG					
Sat_AJ586888	:	GAGGTCAACAACCTTAACAAACG	TATGGGTGATCTGCGTGACATTAACGGTGAGTCCGGTGCA	TGGGCCCGGAATCATGAG				
Pet_AF056581	:	GAGGTCAACAACCTCAACAAACG	TATGGGTGATCTGCGTGACATTAACGGTGAGCCGGTGCA	TGGGCCCGGTATCATGAG				
Sat_AX702523	:	GAGGTCAACAACCTTAACAAACG	TATGGGTGATCTGCGTGACATTAACGGTGAGTCCGGTGCA	TGGGCCCGGAATCATTAG				
SigA_CP000038	:	GAAGTCAACAACCTGAATAAACG	ATGGGGATCTGCGTGACACCAATGGCGGATGCCGGTGCT	TGGGCACGCATAATGAA				
SepA_AF348706	:	GAGGTCAATAAACCTGAACAAAC	GATGGGTGACCTGCGGGATACTCAGGGGATGCCGGTGCT	TGGGCACGCATAATGAA				
SepA_Z48219	:	GAGGTCAATAAACCTGAACAAAC	GATGGGTGACCTGCGGGATACTCAGGGGATGCCGGTGCT	TGGGCCCGCATCATGAA				
EspC_AF297061	:	GAGGTCAATAATCTGAACAAAGC	TATGGGTGACCTGCGGGATACTCAGGGGATGCCGGTGCT	TGGGCCCGCATCATGAA				
EspC_U69128	:	GAGGTCAATAATCTGAACAAAGC	TATGGGTGACCTGCGGGATACTCAGGGGATGCCGGTGCT	TGGGCCCGCATCATGAA				
SepA1_AY604009	:	GAGGTCAACAACCTGAACAAAGC	TATGGGTGACCTGCGTGATACGACGGGGATGCCGGTGCT	TGGGCCCGGTATCATGAA				
EatA_AY163491	:	GAAGTCAATAATCTGAACAAACG	TATGGGAGATTTACGGGATAGTCAGGGAGATGCTG	GAGGGTGCGGTATCATGAA				
Vat_DD002707	:	GAAGTTAACAAACCTGAACAAAC	GATGGGCATTTAGAGGATATTAATGGCGAAGCCGGT	ACGTGGGTGCGTCTGCTGAA				
Hbp_AF218073	:	GAAGTTAACAAACCTGAACAAAC	GATGGGCATTTAGAGGATATTAATGGCGAAGCCGGT	ACGTGGGTGCGTCTGCTGAA				
Boa_AY876285	:	GAAGTCAACAACCTGAACAAAGC	TATGGGGATCTGCGTGATACTCAGGGCGAGGACGGA	ATGTGGGTACGTATCATGAA				
Citr_rodendium	:	GAAGTTAACAAACCTGAACAAAC	GATGGGTGAACCTGCGTGATACGAACCGTGAT	CCCGTGCCTGGGCCGTATCATGAA				
PicS_AE005674	:	GAAGTTAACAAATCTGAACAAAC	GATGGGTGACCTGCGTGACACAAACCGTG	CCCTGGGCCGTATCATGAG				
PicEc_AX276281	:	GAAGTCAACAATCTGAACAAACG	TATGGGTGACCTGCGTGACACAAACCGTG	ATGCCGGAGCCTGGGCCGTATCATGAG				

	*	180	*	200	*	220	*	240	
Espp_X97542	:	ACGGACTGGATTGTTTACCGGTTTCACTGTCAACACACACTGACAGCAGTGCCTCCGCCGATGTTTTCAGTGGTAAAAACG							
Pssa_Y13614	:	ACGGACTGGATTGTTTACCGGTTTCACTGTCAACACACACTGACAGCAGTGCCTCCGCCGATGTTTTCAGTGGTAAAAACG							
Epea_AY258503	:	ACGGACTGGATTGTTTACCGGTTTCACTGTCAACACACACTGACAGCAGTGCCTCCGCCGATGTTTTCAGTGGTAAAAACG							
AJ278144	:	ACGGACTGGATTGTTTACCGGCTTCACTGTCAACACACACTGACAGCAGTGCCTCTGCTGATGCTTTCAAAGGTAAAAACA							
Sat_AJ586888	:	ATGGCCTTGACCTCTTCAACCGGGTGACCATGACCTATACCGACAGCCATGCAGGCAGTGATGCCCTTCAGTGGTGAAAAACG							
Pet_AF056581	:	ATGGCCTTGACCTCTTCAACCGGGTGACCATGACCTATACCGACAGCCATGCAGGCAGTGATGCCCTTCAGTGGTGAAAAACG							
Sat_AX702523	:	ATGGCCTTGACCTCTTCAACCGGGTGACCATGACCTATACCGACAGCCATGCAGGCAGTGATGCCCTTCAGTGGTGAAAAACG							
SigA_CP000038	:	ATGGACTTGACCTTTTCACTGGTCTGACTATGACGTATACCGACAGTCAATGCCAGCAGTAATGTCATTCAGTGGCAAGACG							
Sepa_AF348706	:	ACGGTGTGGATTATTTACGGGGGCAATTGCTGACCTATACGGACAGCAATGCAAGCAGCCACGCATTCAGTGGAAAAAAC							
Sepa_Z48219	:	ACGGTGTGGATTATTTACGGGGGCAATTGCTGACCTATACAGACAGCAATGCAAGCAGCCACGCCTTCAGTGGTAAAAACCC							
Espc_AF297061	:	ACGGTGTGGATTGTTTCAACGGGTGCATTTACTGACCTATACAGACAGCAATGCAAGCAGCCACGCCTTCAGTGGTAAAAACC							
Espc_U69128	:	ACGGTGTGGATTGTTTCAACGGGTGCATTTACTGACCTATACAGACAGCAATGCAAGCAGCCACGCCTTCAGTGGTAAAAACC							
Sepa1_AY604009	:	ATGGTGTGGATTGTTTACCGGTGCATTTGCTGACCTATACGGACAGTAATGCGGACAGTCAATTCAGTGGTAAAAACA							
Eata_AY163491	:	ACGGTATAGATTATTTACCGGTGCATTTACTGACTTATACAGACAAACAATGCTAGCAGCCAGGCTTTCAGCGGTAAAAACA							
Vat_DD002707	:	GAAGTATGGACCTGTTTACCGGGCTGATGGCCACCTACACTGACACAGATGCGTCAGCAGGCTGTACAGCGGTAAAAACA							
Hbp_AF218073	:	GAAGTATGGACCTGTTTACCGGGCTGATGGCCACCTACACTGACACAGATGCGTCAGCAGACCTGTACAGCGGTAAAAACA							
Boa_AY876285	:	CAGGTGCTGACTTGTTCACCTGGTGTGTTGATGAGTTATACCGACAGCAGCCAGTGGAGGGCCTACAGCGCGGACACG							
Citr_rodendum	:	ACGGTGTGGACCTGTTTACCGGGTGTACAGATGACCTATACCGACAGCAGCGCAGACAGTGCCTTCAGCGGGGAAGACA							
Pics_AE005674	:	ACGGTGTGGACCTGTTTACCGGGTGTACAGATGACCTATACCGACAGCAGTGCAGACAGCCATGCATTCAGCGGAAAGACG							
PicEc_AX276281	:	ACGGAGTGGACCTGTTTACCGGTGTCAACGATGACCTATACCGACAGCAGTGCAGACAGCCATGCGTTCAGCGGCAAGACG							
		*	260	*	280	*	300	*	320
Espp_X97542	:	AAGTCTGTGGGGGCTGGCCCTGTATGCTTCCGCCATGTTTGATTCCGGTGCCTATATCGACCTGATTGGCAAAGTATGTTCA							
Pssa_Y13614	:	AAGTCTGTGGGGGCTGGCCCTGTATGCTTCCGCCATGTTTGATTCCGGTGCCTATATCGACCTGATTGGCAAAGTATGTTCA							
Epea_AY258503	:	AAGTCTGTGGGGGCTGGCCCTGTATGCTTCCGCCATGTTTGATTCCGGTGCCTATATCGACCTGATTGGCAAAGTATGTTCA							
AJ278144	:	AAATCTGTGGGGGCCGACCTCTATGCTTCCGCCATGTTTGATTCCGGTGCCTATATCGACCTGATTGGTAAAGTATGTTCA							
Sat_AJ586888	:	AAGTCTGTGGGTGCCGGTCTCTATGCCCTCTGCCATGTTTGAGTCCGGAGCATATATCGACCTCATCGGTAAGTACGTTCA							
Pet_AF056581	:	AAGTCTGTGGGTGCCGGTCTCTATGCCCTCTGCCATGTTTGAGTCCGGAGCATATATCGACCTCATCGGTAAGTACGTTCA							
Sat_AX702523	:	AAGTCTGTGGGTGCCGGTCTCTATGCCCTCTGCCATGTTTGAGTCCGGAGCATATATCGACCTCATCGGTAAGTACGTTCA							
SigA_CP000038	:	AAGTCCGTCGGGGCAGGTCTGTATGCTTCCGCTATATTTGACTCTGCTGCCCTATATCGACCTGATTAGTAAAGTATGTTCA							
Sepa_AF348706	:	AAATCCGTGGGTGCCGGTCTGTATGCCCTCTGCACCTCTTAAATTCGGAGCTTATTTTGACCTGATTGGTAAATATCTCCA							
Sepa_Z48219	:	AAATCCGTGGGTGCCGGTCTGTATGCCCTCTGCACCTCTTAAATTCGGAGCTTATTTTGACCTGATTGGTAAATATCTCCA							
Espc_AF297061	:	AAATCCGTGGGGGAGGGTGTACGCTTCAGCACCTCTTTGATTCCGGGGCTTATTTTGACCTGATTGGTAAATATCTCCA							
Espc_U69128	:	AAATCCGTGGGGGAGGGTGTACGCTTCAGCACCTCTTTGATTCCGGGGCTTATTTTGACCTGATTGGTAAATATCTCCA							
Sepa1_AY604009	:	AAATCAGTGGGTGGTGTGTATGTCATCTGCGCTCTTAAATTCGGGGGCTTATTTTGACCTGATTGGTAAATATCTCCA							
Eata_AY163491	:	AAATCGCTAGGGGAGGGGTGTATGCATCAGGTCTCTTTGAGTCTGGAGCTTATTTTGACCTGATTGGTAAATATCTCCA							
Vat_DD002707	:	AAATCATGGGGTGGTGTTCATGCCAGTGGTCTGTCCGGTCCGGCGCTTACTTTGATTGATTGCCAAATATATTTCA							
Hbp_AF218073	:	AAATCATGGGGTGGTGTTCATGCCAGTGGTCTGTCCGGTCCGGCGCTTACTTTGATGTGATTGCCAAATATATTTCA							
Boa_AY876285	:	CATTGCTCGGGGTGGGATGTACGCATCCGTGATGTTGATTCCGGGGATATATATGGATGTTATCGGCAAGTATATTTCA							
Citr_rodendum	:	AAATCCGTGGGGGCGGCTGTATGCTTCAGCGCTGTTTAAATTCGGCGCTTATATCGATTGATTGGTAAATATATTTCA							
Pics_AE005674	:	AAATCCGTGGGGGCGGCTGTATGCTTCAGCATTTGTTGAGTCCGGTGCCTATATCGATTGATTGGTAAATATATTTCA							
PicEc_AX276281	:	AAATCCGTGGGGGCGGCTGTATGCTTCAGCATTTGTTGAGTCCGGTGCCTATATCGATTGATTGGTAAATATATTTCA							

Figure 3C

EspP_X97542	:	CCATGATAATGAGTACACTGCAACCTTTGCCGGACTCGGAACCCCGTGATTACAGCACGCAATTCATGGTATGCCGGTGCGAG	340	*	360	*	380	*	400
PssA_Y13614	:	CCATGATAATGAGTACACTGCAACCTTTGCCGGACTCGGAACCCCGTGATTACAGCACGCAATTCATGGTATGCCGGTGCGAG							
EpeA_AY258503	:	CCATGATAATGAGTACACTGCAACCTTTGCCGGACTCGGAACCCCGTGATTACAGCACGCAATTCATGGTATGCCGGTGCGAG							
AJ278144	:	TCAATGATAATGAGTACACCGCAACCTTTGCCGGACTCGGAATCCCGTGATTACAGTACGCAATTCATGGTATGCCGGTGCTG							
Sat_AJ586888	:	CCATGACAAACGAGTATACCGCAACTTTTCGCCGGCCCTTGGCACCCAGAGACTACAGCTCCCACTCCTGGTATGCCGGTGCGG							
Pet_AF056581	:	CCATGACAAACGAGTATACCGCAACTTTTCGCCGGCCCTTGGCACCCAGAGACTACAGCTCCCACTCCTGGTATGCCGGTGCGG							
Sat_AX702523	:	CCATGACAAACGAGTATACCGCAACTTTTCGCCGGCCCTTGGCACCCAGAGACTACAGCTCCCACTCCTGGTATGCCGGTGCGG							
SigA_CP000038	:	CCATGATAATGAGTACTCGGCGACCTTTGCTGGACTCGGAACAAAGACTACAGTTCTCATTCCTTGTATGTGGTGCTG							
SepA_AF348706	:	TCAATGATAATCAGCACACGGCGAAATTTTGCCCTCACTGGGAACAAAGACTACAGCTCTCATTCCTGGTATGCCGGTGCTG							
SepA_Z48219	:	TCAATGATAATCAGCACACGGCGAAATTTTGCCCTCACTGGGAACAAAGACTACAGCTCTCATTCCTGGTATGCCGGTGCTG							
EspC_AF297061	:	TCACGACAAATCAGTACACGGCGAGTTTTCGGTCTCTTGGTACAAAGACTACAGCTCTCATTCCTGGTATGCCGGTGCGAG							
EspC_U69128	:	TCACGACAAATCAGTACACGGCGAGTTTTCGGTCTCTTGGTACAAAGACTACAGCTCTCATTCCTGGTATGCCGGTGCGAG							
SepA1_AY604009	:	TCACGATAATCAGTACACGGCGAAATTTTGCCCTCACTGGGAGAAAAGACTACAGCTCTCATTCCTGGTATGCCGGTGCTG							
EatA_AY163491	:	TCAATGATAATCGGTATACGTTGAAATTTTGCCCTCCTTGGGGGAAAGAAAGCTACACCTCCCAATTCCTTGTATGCTGGAGCTG							
Vat_DD002707	:	CAATGAAAACAAATATGACCTGAACTTTGCCGGAGCTGGTAAACAGAACTTCCGCAGCCATTCACTGTATGCAGGTGCAG							
Hbp_AF218073	:	CAATGAAAACAAATATGACCTGAACTTTGCCGGAGCTGGTAAACAGAACTTCCGCAGCCATTCACTGTATGCAGGTGCAG							
Boa_AY876285	:	TCAATGATAATGACTATAACGCCGGTTTTTGCTGGTCTGGGCAACACGGAAATTACGGTACACACTCATGGTATGCTGGCCCTGG							
Citr_rodendium	:	CCATGACAAATGATTACACGGGTAACTTTGCCGGCTTGGGACCGAAGCACTACGGAAACCCACTCCTGGTATGCCGGAGCGG							
PicS_AE005674	:	CCATGACAAATGATTACACAGGTAACTTTGCTAGCTGGGAACGAAACACTACAACACCCATTCTCTGGTATGCCGGTGCTG							
PicEc_AX276281	:	CCATGACAAATGATTACACAGGTAACTTTGCCGGTCTGGGAACGAAACACTACAACACCCATTCTCTGGTATGCCGGTGCTG							
EspP_X97542	:	AAGCGGGCTACCGCTATCATGTCACTGAGGATGCCTGGATTGAGCCACAGGCTGAGCTGGTTTACGGTTCCTGTATCCGGT	420	*	440	*	460	*	480
PssA_Y13614	:	AAGCGGGCTACCGCTATCATGTCACTGAGGATGCCCTGGATTGAGCCACAGGCTGAGCTGGTTTACGGTTCCTGTATCCGGT							
EpeA_AY258503	:	AAGTGGGCTACCGCTATCATGTCACTGAGGATGCCCTGGATTGAGCCACAGGCTGAGCTGGTTTACGGTTCCTGTATCCGGT							
AJ278144	:	AAGCAGGCTACCGCTGTATGTCACTGAGGATACCTGGATTGAGCCACAGGCAAACTGGTTTACGGTTCCTGTATCCGGT							
Sat_AJ586888	:	AAGTCGGTTACCGTTACCATGTAACTGACTCTGCAATGGATTGAGCCGCAGGCGAACTTGTTCAGGTCCTGTATCCGGG							
Pet_AF056581	:	AAGTCGGTTACCGTTACCATGTAACTGACTCTGCAATGGATTGAGCCGCAGGCGAACTTGTTCAGGTCCTGTATCCGGG							
Sat_AX702523	:	AAGTCGGTTACCGTTACCATGTAACTGACTCTGCAATGGATTGAGCCGCAGGCGAACTTGTTCAGGTCCTGTATCCGGG							
SigA_CP000038	:	AAGCAGGCTACCGCTATCATGTAAACAGAAAGACTCCTGGATTGAGCCGCAGGCAAACTGGTTTATGGGCCGTATCAGGT							
SepA_AF348706	:	AAGTTGGTTATCGTTACCACTGACGAAAGAGTCCCTGGTGGAGCCACAGATAGAGCTGGTTTACGGTTCCTGTATCAGGA							
SepA_Z48219	:	AAGTTGGTTATCGTTACCACTGACGAAAGAGTCCCTGGTGGAGCCACAGATAGAGCTGGTTTACGGTTCCTGTATCAGGA							
EspC_AF297061	:	AGTCCGGTATCGTTACCACTGTGGAAGAGTCCCTGGTGGAGCCACAGATGGAGCTGGTTTACGGTTCCTGTGTACAGGA							
EspC_U69128	:	AGTCCGGTATCGTTACCACTGTGGAAGAGTCCCTGGTGGAGCCACAGATGGAGCTGGTTTACGGTTCCTGTGTACAGGA							
SepA1_AY604009	:	AAGTCGGGTATCGTTACCACTGTGGAAGAAATCCTGGTGGAACTCAGATGGAGCTGGTTTACGGTTCCTGTTTCAGGA							
EatA_AY163491	:	AAATCCGGGTATCGTTATCACATGTCAGAAAATACATGGGTGGAAACCCACAGATGGAACCTGGTTTATGGTTTCGGTATCAGGA							
Vat_DD002707	:	AAGTCGGATACCGTTATCATCTGACAGATACGAGTTTGTGAACCTCAGCGGAACTGGTCTGGGGAAGACTGCAGGGC							
Hbp_AF218073	:	AAGTCGGATACCGTTATCATCTGACAGATACGACGTTTGTGAACCTCAGGCGGAACTGGTCTGGGGAAGACTGCAGGGC							
Boa_AY876285	:	AAGCGGGATACCGTTACCGTCTGACAGAAAAGCCCTGTATATTGAGCCGCAGGCGGAACTGGTATATGGAACCGTCTCCGGA							
Citr_rodendium	:	AAACGGGTACCGTTATCACCTGACGGAGGACACCTTTATTGAGCCGCAGGCGGAACTGGTTTATGGCGCGGTGTCGGC							
PicS_AE005674	:	AAACGGGTACCGCTATCACCTGACAGAGGACACGTTCAATTGAGCCGCAGGCTGAACCTGGTTTACGGCGCGGTGTCGGG							
PicEc_AX276281	:	AAACGGGTACCGCTATCACCTGACAGAGGAAACGTTCAATTGAGCCGCAGGCTGAACCTGGTTTACGGCGCGGTGTCGGG							

Figure 3D	*	500	*	520	*	540	*	560
EspP_X97542	:	AAACAGTTTGCATGGGAAGGACCAAGGAAATGCATCTGTCCATGAAGGACAAAGGACTACAAATCCGCTGATTGGCCCGAACCGG						
Pssa_Y13614	:	AAACAGTTTGCATGGGAAGGACCAAGGAAATGCATCTGTCCATGAAGGACAAAGGACTACAAATCCGCTGATTGGCCCGAACCGG						
EpeA_AY258503	:	AAACAGTTTGCATGGGAAGGACCAAGGAAATGCATCTGTCCATGAAGGACAAAGGACTACAAATCCGCTGATTGGCCCGAACCTGG						
AJ278144	:	AAACAGTTTGCATGGGAAGGACCAAGGAAATGCATCTGTCCATGAAGGACAAAGGACTACAAATCCGCTGATTGGTCGTACCCGG						
Sat_AJ586888	:	AAACAGTTTCTCCTGGGAAGGACCAAGGAAATGAACCTCACCATGAAGGATAAAGGACTTTAATCCGCTGATTGGGCGTACCCGG						
Pet_AF056581	:	AAACAGTTTCTCCTGGGAAGGACCAAGGAAATGAACCTCACCATGAAGGATAAAGGACTTTAATCCGCTGATTGGGCGTACCCGG						
Sat_AX702523	:	AAACAGTTTCTCCTGGGAAGGACCAAGGAAATGAACCTCACCATGAAGGATAAAGGACTTTAATCCGCTGATTGGGCGTACCCGG						
SigA_CP000038	:	AAACGGTTCGACTGGCAGGATCGCGGAAATGAGCGTGACCATGAAGGATAAAGGACTTTAATCCGCTGATTGGGCGTACCCGG						
SepA_AF348706	:	AAAGCTTTTAGCTGGGAAGACCGGGGAAATGGCTCTGAGCATGAAGACAAAGGATTATAAACCCTGATTGGCCCGTACTGG						
SepA_Z48219	:	AAAGCTTTTAGCTGGGAAGCCCGGGGAAATGGCTCTGAGCATGAAGACAAAGGATTATAAACCCTGATTGGCCCGTACTGG						
EspC_AF297061	:	AAATCTTTTAGCTGGGAAGACCGGGGAAATGGCCCTGAGCATGAAGACAAAGGATTATAAACCCTGATTGGCCCGTACCCGG						
EspC_U69128	:	AAATCTTTTAGCTGGGAAGACCGGGGAAATGGCCCTGAGCATGAAGACAAAGGATTATAAACCCTGATTGGCCCGTACCCGG						
SepA1_AY604009	:	AAATCCCTTTAACTGGGAAGACCGGGGAAATGGCACTGACCATGAAGACAAAGGATTATAAACCCTGATTGGTCGTACCCGG						
EatA_AY163491	:	AAGTCATTTAACTGGAAAGACCAAGGAAATGCAACTGAGTATGAAGACAAAGACTATCACCCCTGTAATTCCTGTTAGGCAGAACAGG						
Vat_DD002707	:	CAAAACATTTAACTGGGAACGACAGTGGAAATGGATGTCTCAATGCGTCGTAACAGCGTTAATCCTCTGTTAGGCAGAACCCGG						
Hbp_AF218073	:	CAAAACATTTAACTGGGAACGACAGTGGAAATGGATGTCTCAATGCGTCGTAACAGCGTTAATCCTCTGTTAGGCAGAACCCGG						
Boa_AY876285	:	ACAACGCTGAAATGGAAATGATAATGGTATGGATGTGTCGATGCCAGCAAAACGTAATAATCCGTTGATAGGCGGTACAGG						
Citr_rodendium	:	AAAACGTTCCGCTGGAAAGACGCGTGATATGGAACCTGAGCATGAAGACAAAGGATTTTCAGCCCCGTTGATTGGCAGAACACAGG						
Pics_AE005674	:	AAAACATTCGCTGGAAAGACGCGTGATATGGAACCTGAGCATGAAGACAAAGGACTTCAGTCCGCTGGTTGGAAAGAACACAGG						
PicEc_AX276281	:	AAAACATTCGCTGGAAAGACGCGTGATATGGAACCTGAGCATGAAGACAAAGGACTTCAGTCCGCTGATTGGAAAGAACACAGG						
EspP_X97542	:	TGTGGATGTGGGTAAATCCTTCTCTGGTAAGGACTGGAAAGTGACAGCCCCGTGCCCGG	*	580	*	600	*	
Pssa_Y13614	:	TGTGGATGTGGGTAAATCCTTCTCTGGTAAGGACTGGAAAGTGACAGCCCCGTGCCCGG						
EpeA_AY258503	:	TGTGGATGTGGGTAAATCCTTCTCTGGTAAGGACTGGAAAGTGACAGCCCCGTGCCCGG						
AJ278144	:	TGTGGATGTGGGTAAATCCTTCTCAAGTAAGGACTGGAAAGTGACAGCCCCGTGCCCGG						
Sat_AJ586888	:	TGTTGATGTGGGTAAATCCTTCTCGGGTAAGGACTGGAAAGTCAAGCCCCGCGCCCGG						
Pet_AF056581	:	TGTTGATGTGGGTAAATCCTTCTCGGGTAAGGACTGGAAAGTCAAGCCCCGCGCCCGG						
Sat_AX702523	:	TGTTGATGTGGGTAAATCCTTCTCGGGTAAGGACTGGAAAGTCAAGCCCCGCGCCCGG						
SigA_CP000038	:	TGTTGATGTGGGTAAATCCTTCTCGGGTAAGGACTGGAAAGTCAAGCCCCGCGCCCGG						
SepA_AF348706	:	TGTTGACGTGGGAAGAGCCCTTCTCCGGTAAGGACTGGAAAGTCAAGCCCCGCGCCCGG						
SepA_Z48219	:	TGTTGACGTGGGAAGAGCCCTTCTCCGGAGACGACTGGAAATCACAGCTCGAGCCCGG						
EspC_AF297061	:	TGTTGACGTGGGAAGAACCTTCTCCGGAGACGACTGGAAATTACCCGCGGAGCCCGG						
EspC_U69128	:	TGTTGACGTGGGAAGAACCTTCTCCGGAGACGACTGGAAATTACCCGCGGAGCCCGG						
SepA1_AY604009	:	TGTTGACGTGGGAAGAAAGCTTCTCCGGAGATGACTGGAAATCACAGCGCGCCCGCGG						
EatA_AY163491	:	TGTGGATGTAGGTAGAGCGTTCTCTGGAGATACCTGGAAAGTAACAGTACGTGCAGG						
Vat_DD002707	:	CGTTGTTTCCGGTAAACCTTTCAGTGGTAAGGACTGGAGTCTGACAGCCCCGTGCCCGG						
Hbp_AF218073	:	CGTTGTTTCCGGTAAACCTTTCAGTGGTAAGGACTGGAGTCTGACAGCCCCGTGCCCGG						
Boa_AY876285	:	CGTGGCATTTGGGCAAAACGTTTCAGTGGCAAAGGACTGGAGCGTTACGGCCCCGTACAGG						
Citr_rodendium	:	GATTGAGCTGGGTAAACCTTTCAGTGGTAAGGACTGGCGCGTACGGCCCCGGCCCGG						
Pics_AE005674	:	GGTTGAAC TGGGCAAGACCTTTCAGTGGTAAGGACTGGAGTGTGACGGCCCCGTGCCCGG						
PicEc_AX276281	:	GATTGAAC TGGGCAAGACCTTTCAGTGGTAAGGACTGGAGTGTGACGGCCCCGTGCCCGG						

Figure 4A Alignment of pic-like gene with other SPATE genes.

	*	20	*	40	*	60	*	80
NZ_AAJV01000028_Pic-like_	:	GAAGTTAACAAACCTGAACAAACG	TATGGGTGAAC	TGCGTGATACGAAACGGT	GATGCCGGTGCC	TGGCTGGCCCGTATCATGAA		
citrobacter_freundii_	:	GAAGTTAACAAACCTGAACAAACG	TATGGGTGAAC	TGCGTGATACGAAACGGT	GATGCCGGTGCC	TGGCTGGCCCGTATCATGAA		
AE005674_Pic_Shigella_	:	GAAGTTAACAAATCTGAACAAACG	TATGGGTGACCT	GTGACACAAACGGT	GATGCCGGTGCC	TGGCTGGCCCGTATCATGAG		
AX276281_Pic_E.coli_	:	GAAGTCAACAATCTGAACAAACG	TATGGGTGACCT	GTGACACAAACGGT	GATGCCGGTGCC	TGGCTGGCCCGTATCATGAG		
X97542_EspP_	:	GAGGTCAACAACCTGAACAAACG	TATGGGTGACCT	GTGATATCAACGGCGAAGCC	GGTGATGGGCACGCATCATGAG			
Y13614_PssA_	:	GAGGTCAACAACCTGAACAAACG	TATGGGTGACCT	GTGATATCAACGGCGAAGCC	GGTGATGGGCACGCATCATGAG			
AY258503_EpeA_	:	GAGTCAACAACCTGAACAAACG	TATGGGTGACCT	GTGACATCAATGGCGAAGCC	GGTGATGGGCACGCATCATGAG			
AJ278144_EspI_	:	GAGGTCAACAACCTGAATAAACG	TATGGGTGACCT	GTGACATCAATGGCGAAGCC	GGTGATGGGCACGTATCATGAG			
AF056581_Pet_	:	GAGGTCAACAACCTGAATAAACG	TATGGGTGACCT	GTGACATCAATGGCGAAGCC	GGTGATGGGCACGTATCATGAG			
AJ586888_Sat_	:	GAGGTCAACAACCTGAATAAACG	TATGGGTGACCT	GTGACATCAATGGCGAAGCC	GGTGATGGGCACGTATCATGAG			
CP000038_SigA_	:	GAGGTCAACAACCTGAATAAACG	TATGGGTGACCT	GTGACATCAATGGCGAAGCC	GGTGATGGGCACGTATCATGAG			
AF297061_EspC1_	:	GAGGTCAACAACCTGAATAAACG	TATGGGTGACCT	GTGACATCAATGGCGAAGCC	GGTGATGGGCACGTATCATGAG			
U69128_EspC2_	:	GAGGTCAACAACCTGAATAAACG	TATGGGTGACCT	GTGACATCAATGGCGAAGCC	GGTGATGGGCACGTATCATGAG			
Z48219_SepA_	:	GAGGTCAACAACCTGAATAAACG	TATGGGTGACCT	GTGACATCAATGGCGAAGCC	GGTGATGGGCACGTATCATGAG			
AY604009_SepA1_	:	GAGGTCAACAACCTGAATAAACG	TATGGGTGACCT	GTGACATCAATGGCGAAGCC	GGTGATGGGCACGTATCATGAG			
AY163491_EatA_	:	GAGGTCAACAACCTGAATAAACG	TATGGGTGACCT	GTGACATCAATGGCGAAGCC	GGTGATGGGCACGTATCATGAG			
DD002707_Vat_	:	GAGGTCAACAACCTGAATAAACG	TATGGGTGACCT	GTGACATCAATGGCGAAGCC	GGTGATGGGCACGTATCATGAG			
AF218073_Hbp_	:	GAGGTCAACAACCTGAATAAACG	TATGGGTGACCT	GTGACATCAATGGCGAAGCC	GGTGATGGGCACGTATCATGAG			
AY876285_Boa_	:	GAGGTCAACAACCTGAATAAACG	TATGGGTGACCT	GTGACATCAATGGCGAAGCC	GGTGATGGGCACGTATCATGAG			
NZ_AAJV01000028_Pic-like_	:	CGGCGCAGGTTTCAGCCCGATGG	CGGGGTACAGTGATAATTACACT	CACGTTTCAGGTCGGCTTTGAC	AAAAAACAATGTGCTGG			
citrobacter_freundii_	:	CGGCGCCGATTCAGCAGAGGGCG	GTACAGTGATAACTACACCCACG	TTCAGGTCGGTTTGACAAAAACA	TGCGCTGG			
AE005674_Pic_Shigella_	:	TGGTGCCGGTCTGCAGACGGTGT	TACAGTGATAATTACACCCATG	TTCAGGTCGGCTTTGACAAAAACA	TGAACTGG			
AX276281_Pic_E.coli_	:	TGGTGCCGGTCTGCAGACGGTGT	TACAGTGATAATTACACCCATG	TTCAGGTCGGCTTTGACAAAAACA	TGAACTGG			
X97542_EspP_	:	CGGTACCCGGCTCTGCCAGTGGT	GTTCAGTGACAACTACACGCACG	TTCAGGTCGGGTTCGACAAAAACA	CACGAGCTGG			
Y13614_PssA_	:	CGGTACCCGGCTCTGCCAGTGGT	GTTCAGTGACAACTACACGCACG	TTCAGGTCGGGTTCGACAAAAACA	CACGAGCTGG			
AY258503_EpeA_	:	CGGTACTGGCTCTGCCAGTGGT	GTTCAGTGACAACTACACGCACG	TTCAGGTCGGGTTCGACAAAAACA	TGAGCTGG			
AJ278144_EspI_	:	CGGAACCCGGCTCTGCCGGTGG	TTCAGTGACAACTACACCCACG	TTCAGGTCGGGTTCGACAAAAACA	TGAGCTGG			
AF056581_Pet_	:	TGGAACCCGGTCTGCCGGCGGT	GGATTTCAGTGACAACTACACCCACG	TTCAGGTCGGGTTCGACAAAAACA	TGAACTCG			
AJ586888_Sat_	:	CGGAACCCGGTCTGCCGGCGGT	GGATTTCAGTGACAACTACACCCACG	TTCAGGTCGGGTTCGACAAAAACA	TGAACTCG			
CP000038_SigA_	:	CGGAGCAGGTTTCAGCTTCTAG	TGGATACAGTGACAACTACACCCAT	GTGCAGATTGGTGTGGATAAA	AAACAATGAGCTGG			
AF297061_EspC1_	:	CGGTACCCGGTTCGGCAGATGG	TGGTTACAGCGATAACTACACT	CACGTTTCAGATTGGTCCGAC	AGAAAGCATGAGCTGG			
U69128_EspC2_	:	CGGTACCCGGTTCGGCAGATGG	TGGTTACAGCGATAACTACACT	CACGTTTCAGATTGGTCCGAC	AGAAAGCATGAGCTGG			
Z48219_SepA_	:	TGGTACCCGGTTCGGCAGATGG	TGACTACAGCGATAACTACACT	CACGTTTCAGATTGGTCCGAC	AGAAAGCATGAGCTGG			
AY604009_SepA1_	:	CGGTACCCGGTTCGGCAGATGG	TGACTACAGCGATAACTACACT	CACGTTTCAGATTGGTCCGAC	AGAAAGCATGAGCTGG			
AY163491_EatA_	:	TGGTACCCGGTTCAGGTGAGAG	TGGTTACAGAGATAACTATACC	CCACGTTTCAGATTGGTGCAG	ACAGAAAGCATGAGCTGA			
DD002707_Vat_	:	CGGTTCCGGCTCTGCTGATGG	CGGGTTCACCTGCTCAGATTG	CGGGGCTGACCCGTAAAGC	ACGAACCTGG			
AF218073_Hbp_	:	CGGTTCCGGCTCTGCTGATGG	CGGGTTCACCTGCTCAGATTG	CGGGGCTGACCCGTAAAGC	ACGAACCTGG			
AY876285_Boa_	:	CGGCGCCGGAACCCGGTGAC	CGCCGGATATTCTGATCGTT	ACACCCATCTGCAACCGGG	TTTGATAAAAAACA	CCGGTTGT		

	*	180	*	200	*	220	*	240
ANZ_AAJV01000028_pic-like_	:	ATGGCGTGGACCTGTTTACCGGGGATCACAAATGACCTATACCGACAGCAGTGCAGACAGTGATGCGTT	CAGCGG	AAAAACA				
citrobacter_freundii_	:	ACGGTGTGGACCTGTTTACCGGTGTCAAGATGACCTATACCGACAGCAGCGCAGACAGTGATGCGTT	CAGCGG	AAAAACA				
AE005674_Pic_Shigella_	:	ACGGTGTGGACCTGTTTACCGGTGTCAAGATGACCTATACCGACAGCAGTGCAGACAGCATTCAGCGG	AAAAACA					
AX276281_Pic_E.coli_	:	ACGGAGTGGACCTGTTTACCGGTGTCAAGATGACCTATACCGACAGCAGTGCAGACAGCATGCGTT	CAGCGG	AAAAACA				
X97542_EspP_	:	ACGGACTGGATTGTTTACCGGTTTCACTGTCACACACACTGACAGCAGTGCCCTCCGCCGATGTTT	CAGTGGT	AAAAACG				
Y13614_PssA_	:	ACGGACTGGATTGTTTACCGGTTTCACTGTCACACACACTGACAGCAGTGCCCTCCGCCGATGTTT	CAGTGGT	AAAAACG				
AY258503_EpeA_	:	ACGGACTGGATTGTTTACCGGTTTCACTGTCACACACACTGACAGCAGTGCCCTCCGCCGATGTTT	CAGTGGT	AAAAACG				
AJ278144_EspI_	:	ACGGACTGGATTGTTTACCGGCTTCACTGTCACACACACTGACAGCAGTGCCCTCTGCTGATGCTT	TCAAAGG	TA				
AF056581_Pet_	:	ATGGCCCTTGACCTCTTCAACCGGGTGACCATGACCTATACCGACAGCCATGACGAGTGATGCC	TTT	CAGTGGT	AAAAACG			
AJ586888_Sat_	:	ATGGCCCTTGACCTCTTCAACCGGGTGACCATGACCTATACCGACAGCCATGACGAGTGATGCC	TTT	CAGTGGT	AAAAACG			
CP000038_SigA_	:	ATGGACTTGACCTTTTCACTGGTCTGACTATGACGTAATACCGACAGTCAATGCCAGCAGTAA	TGCAAT	TTCAGTGGC	AAAAACG			
AF297061_EspC1_	:	ACGGTGTGGATTGTTTCAACGGGTGCATTACTGACCTATACAGACAGCAATGCAAGCAGCCAC	GCCTT	CAGTGGT	AAAAAC			
U69128_EspC2_	:	ACGGTGTGGATTGTTTCAACGGGTGCATTACTGACCTATACAGACAGCAATGCAAGCAGCCAC	GCCTT	CAGTGGT	AAAAAC			
Z48219_SepA_	:	ACGGTGTGGATTATTTACGGGGGCAATGCTGACCTATACGGACAGCAATGCCAAGCAGCCAC	GCATTC	CAGTGGAAAAA	AC			
AY604009_SepA1_	:	ATGGTGTGGATTGTTTACCGGTGCATTGCTGACCTATACGGACAGTAATGCCGACAGTCATG	CATTC	CAGTGGT	AAAAACA			
AY163491_EatA_	:	ACGGTATAGATTATTCACCCGGTGCAATTA	CTGACTT	TACAGACA	CAATGCTAGCAGCCAGGC	TTT	CAGCGG	TA
DD002707_Vat_	:	GAAATA	TGGACCTGTTTACCGGCGTGATGGCCACCTACACTGACACAGATGCGT	CAGCAGG	CCTGTACAGCGG	TA	AAAAACA	
AF218073_Hbp_	:	GAAATA	TGGACCTGTTTACCGGCGTGATGGCCACCTACACTGACACAGATGCGT	CAGCAGC	CTGTACAGCGG	TA	AAAAACA	
AY876285_Boa_	:	CAGGTGCTGACTTGTTC	ACTGGTGTGTTGATGAGTTATACCGACAGCAGCGCCAGTGGACGG	CCTAC	CAGCGG	CG	AC	

	*	260	*	280	*	300	*	320
NZ_AAJV01000028_Pic-like_	:	AAATCCGTGGGGGGGGTCTGTATGCTTCAGCATTTGTTAACTCCGGTGCCCTACATTGATTTGATTGGTAAATACATTCA						
citrobacter_freundii_	:	AAATCCGTGGGGGGGGCTGTATGCTTCAGCGCTGTTTAAATCCGGCGCCTATATCGATTGATTTGTTGTTAAATATATTCA						
AE005674_Pic_Shigella_	:	AAATCCGTGGGGGGGGTCTGTATGCTTCAGCATTTGTTGAGTCCGGTGCCCTATATCGATTGATTTGTTGTTAAATATATTCA						
AX276281_Pic_E.coli_	:	AAATCCGTGGGGGGGGTCTGTATGCTTCAGCATTTGTTGAGTCCGGTGCCCTATATCGATTGATTTGTTGTTAAATATATTCA						
X97542_EspP_	:	AAGTCTGTGGGGGCTGGCCTGTATGCTTCCGCCATGTTTGATTCCGGTGCCCTATATCGACCTGATTTGGCAAGTATGTTTCA						
Y13614_PssA_	:	AAGTCTGTGGGGGCTGGCCTGTATGCTTCCGCCATGTTTGATTCCGGTGCCCTATATCGACCTGATTTGGCAAGTATGTTTCA						
AY258503_EpeA_	:	AAGTCTGTGGGGGCTGGCCTGTATGCTTCCGCCATGTTTGATTCCGGTGCCCTATATCGACCTGATTTGGCAAGTATGTTTCA						
AJ278144_EspI_	:	AAATCTGTGGGGCCGGACTCTATGCTTCCGCCATGTTTGATTCCGGTGCCCTATATCGACCTGATTTGGTAAAGTATGTTTCA						
AF056581_Pet_	:	AAGTCTGTGGGTGCCGGTCTCTATGCCCTCTGCCATGTTTGAGTCCGGAGCATATATCGACCTCATCGGTAAGTACGTTTCA						
AJ586888_Sat_	:	AAGTCTGTGGGTGCCGGTCTCTATGCCCTCTGCCATGTTTGAGTCCGGAGCATATATCGACCTCATCGGTAAGTACGTTTCA						
CP000038_SigA_	:	AAGTCCGTGGGGCAGGTCTGTATGCTTCCGCTATATTTGACTCTGGTGCCCTATATCGACCTGATTTAGTAAAGTATGTTTCA						
AF297061_EspC1_	:	AAATCCGTGGGGGGAGGGTTGTACGCTTCAGCACTCTTTGATTTCCGGGGCTTATTTTGACCTGATTTGGTAAATATCTCCCA						
U69128_EspC2_	:	AAATCCGTGGGGGGAGGGTTGTACGCTTCAGCACTCTTTGATTTCCGGGGCTTATTTTGACCTGATTTGGTAAATATCTCCCA						
Z48219_SepA_	:	AAATCAGTGGGTGGTCTGTATGCCCTCTGCACCTCTTTAATTCGGGAGCTTATTTTGACCTGATTTGGTAAATATCTCCCA						
AY604009_SepA1_	:	AAATCCGTAGGGGGAGGGGTGTATGCATCAGGTCCTCTTTGAGTCTGGAGCTTATTTTGACCTGATTTGGTAAATATCTCCCA						
AY163491_EatA_	:	AAATCATGGGGTGGTGGTTTCTATGCCAGTGGTCTGTTCCGGTCCGGCGCTTACTTTGATTTGATTTGCCAAATATATTCA						
DD002707_Vat_	:	AAATCATGGGGTGGTGGTTTCTATGCCAGTGGTCTGTTCCGGTCCGGCGCTTACTTTGATTTGATTTGCCAAATATATTCA						
AF218073_Hbp_	:	CAATTCGCTCGGGGGTGGGATGTACGCATCCGCTGATGTTTGATTCGGGGGATATATATGGATGTTATCGGCAAGTATATTCA						
AY876285_Boa_	:							

Figure 4C

	*	340	*	360	*	380	*	400
NZ_AAJV01000028_Pic-like_	:	CCATAACAATGATTACACAGGCAACTTTGCCGGTCTGGGTACGAAGCACTACGGAACCCACTCCTGGTATGCCGGAGCGG						
citrobacter_freundii_	:	CCATGACAAATGATTACACGGGTAACTTTGCCGGCTGGGACGAAGCACTACGGAACCCACTCCTGGTATGCCGGAGCGG						
AE005674_Pic_Shigella_	:	CCATGACAAATGATTACACAGGTAACCTTTGCTAGCCTGGGAACGAACACTAACACCCATCCTGGTATGCCGGTGCTG						
AX276281_Pic_E.coli_	:	CCATGACAAATGATTACACAGGTAACCTTTGCCGGTCTGGGAACGAACACTAACACCCATCCTGGTATGCCGGTGCTG						
X97542_Essp_	:	CCATGATAATGAGTACACTGCAACCTTTGCCGGACTCGGAACCCGTGATTACAGCACGCATTCATGGTATGCCGGTGCA						
Y13614_Pssa_	:	CCATGATAATGAGTACACTGCAACCTTTGCCGGACTCGGAACCCGTGATTACAGCACGCATTCATGGTATGCCGGTGCA						
AY258503_EpeA_	:	CCATGATAATGAGTACACTGCAACCTTTGCCGGACTCGGAACCCGTGATTACAGCACGCATTCATGGTATGCCGGTGCA						
AJ278144_EspI_	:	TCATGATAATGAGTACACTGCAACCTTTGCCGGACTCGGAATCCGTGATTACAGTACGCATTCATGGTATGCCGGTGCTG						
AF056581_Pet_	:	CCATGACAAACGAGTATACCGCAACTTTGCCGGCTTGGCACAGAGACTACAGTCCCACCTCCCTGGTATGCCGGTGCGG						
AJ586888_Sat_	:	CCATGACAAACGAGTATACCGCAACTTTGCCGGCTTGGCACAGAGACTACAGTCCCACCTCCCTGGTATGCCGGTGCGG						
CP000038_SigA_	:	CCATGATAATGAGTACTCGGCGACCTTTGCTGGAACAAAAGACTACAGTTCTCATTCCTTGTATGTGGTGCTG						
AF297061_EspC1_	:	TCACGACAAATCAGTACACGGCGAGTTTTCGCTCTTGGTACAAAAGACTACAGTCTCATTCCTGGTATGCCGGTGCA						
U69128_EspC2_	:	TCACGACAAATCAGTACACGGCGAGTTTTCGCTCTTTCGCTCTTGGTACAAAAGACTACAGTCTCATTCCTGGTATGCCGGTGCA						
Z48219_SepA_	:	TCATGATAATCAGCACACGGCGAATTTTGCCCTCACTGGGAACAAAAGACTACAGTCTCATTCCTGGTATGCCGGTGCTG						
AY604009_SepA1_	:	TCACGATAATCAGTACACGGCGAATTTTGCCCTCACTGGAGAAAAGACTACAGTCTCATTCCTGGTATGCCGGTGCTG						
AY163491_EatA_	:	TCATGATAATTCGGGTATACGTTGAAATTTTCGCTCCTTGGGGGAAAGAAAGCTACACCTCCCATTCCTTGTATGCTGGAGCTG						
DD002707_Vat_	:	CAATGAAAACAAAATATGACCTGAACCTTTGCCGGAGCTGGTAAACAGAACTTCGCGAGCCATTCATCTGTATGCAGGTGCAG						
AF218073_Hbp_	:	CAATGAAAACAAAATATGACCTGAACCTTTGCCGGAGCTGGTAAACAGAACTTCGCGAGCCATTCATCTGTATGCAGGTGCAG						
AY876285_Boa_	:	TCATGATAATGACTATAACGCCGGTTTTCGCTGGTCTGGGCAACCGGAATTACGGTACACACTCATGGTATGCTGGCCTGG						
	*	420	*	440	*	460	*	480
NZ_AAJV01000028_Pic-like_	:	AAACGGGTTACCGTTATCACCTGACGGGAAGACACATTTATTGAGCCTCAGGCCCGAACTGGTTTACGGCGCAGTGTC						
citrobacter_freundii_	:	AAACGGGTTACCGTTATCACCTGACGGAGGACACCTTTATTGAGCCGAGGCCGAACTGGTTTATGCGCGGTGTCCGGC						
AE005674_Pic_Shigella_	:	AAACGGGTTACCGCTATCACCTGACAGAGGACACGTTTCATTGAGCCGCAAGCTGAACCTGGTTTACGGCGCCGTGTCCGGG						
AX276281_Pic_E.coli_	:	AAACGGGTTACCGCTATCACCTGACAGAGGAAACGTTTCATTGAGCCGCAAGCTGAACCTGGTTTACGGCGCCGTGTCCGGG						
X97542_Essp_	:	AAGCGGGCTACCGCTATCATGTCACTGAGGATGCCCTGGATTGAGCCACAGGCTGAGCTGGTTTACGGTTCTGTATCCGGT						
Y13614_Pssa_	:	AAGCGGGCTACCGCTATCATGTCACTGAGGATGCCCTGGATTGAGCCACAGGCTGAGCTGGTTTACGGTTCTGTATCCGGT						
AY258503_EpeA_	:	AAGTGGGCTACCGCTATCATGTCACTGAGGATGCCCTGGATTGAGCCACAGGCTGAGCTGGTTTACGGTTCTGTATCCGGT						
AJ278144_EspI_	:	AAGCAGGCTACCGCTGTCACTGAGGATACCTGGATTGAGCCACAGGCAGAACTGGTTTACGGTGTGTATCCGGT						
AF056581_Pet_	:	AAGTCGGTTACCGTTACCATGTAACTGACTCTGCAATGGATTGAGCCGCAAGCCGAACTTGTTCACGGTGTGTATCCGGG						
AJ586888_Sat_	:	AAGTCGGTTACCGTTACCATGTAACTGACTCTGCAATGGATTGAGCCGCAAGCCGAACTTGTTCACGGTGTGTATCCGGG						
CP000038_SigA_	:	AAGCAGGCTACCGCTATCATGTAAACAGAAAGACTCCTGGATTGAGCCGCAAGCCGAACTGGTTTATGGGCGCCGTATCAGGT						
AF297061_EspC1_	:	AGGTCCGGGTATCGTTACCACTGTGGAAGAGTCCCTGGGTGGAGCCACAGATGGAGCTGGTTTACGGTTCTGTGTACAGGA						
U69128_EspC2_	:	AGGTCCGGGTATCGTTACCACTGTGGAAGAGTCCCTGGGTGGAGCCACAGATGGAGCTGGTTTACGGTTCTGTGTACAGGA						
Z48219_SepA_	:	AAGTTGGTTATCGTTACCACTGACGAAAGAGTCCCTGGGTGGAGCCACAGATAGAGCTGGTTTACGGTTCTGTATCAGGA						
AY604009_SepA1_	:	AAGTCGGGTATCGTTACCACTGTGGAAGAAATCCTGGGTGGAACTCAGATGGAGCTGGTTTACGGTTCTGTTCAGGA						
AY163491_EatA_	:	AAATCCGGGTATCGTTATCACATGTCAGAAATAACATGGGTGGAACTCAGATGGAACTGGTTTATGGTTCGGTATCAGGA						
DD002707_Vat_	:	AAGTCGGGATACCGTTATCATCTGACAGATACGACGTTTGTGTAACCTCAGGCGGAACTGGTCTGGGAAAGACTGCAGGGC						
AF218073_Hbp_	:	AAGTCGGGATACCGTTATCATCTGACAGATACGACGTTTGTGTAACCTCAGGCGGAACTGGTCTGGGAAAGACTGCAGGGC						
AY876285_Boa_	:	AAGGCGGATACCGTTACCGTCTGACAGAAAGCCGTATATTGAGCCGCAAGCCGGAACCTGTTATGGAAACCCGTCTCCGGA						

Figure 4D

NZ_AAJV01000028_Pic-like_	:	AAAACATTC	CGCTG	GAAAG	ACGGT	GAGAT	GGACCT	GAGTAT	GAAACA	AAGGAT	TTCAG	CCCGT	TGATT	TGGC	AGAAC	AGG
citrobacter_freundii_	:	AAAACGT	TCCGCT	GGAAG	ACGGT	GATAT	GGACCT	GAGCAT	GGAACA	AAGGAT	TTCAG	CCCGT	TGATT	TGGC	AGAAC	AGG
AE005674_Pic_Shigella_	:	AAAACAT	TCCGCT	GGAAG	ACGGT	GATAT	GGACCT	GAGCAT	GGAACA	AGGACT	TTCAG	TCCGCT	GTG	TGGA	AGAAC	AGG
AX276281_Pic_E.coli_	:	AAAACAT	TCCGCT	GGAAG	ACGGT	GATAT	GGACCT	GAGCAT	GGAACA	AGGACT	TTCAG	TCCGCT	GTG	TGGA	AGAAC	AGG
X97542_EspP_	:	AAACAGT	TTGCAT	GGAAG	ACG	GGAAT	GCAT	CTGT	CCAT	GAA	GAC	TAC	AA	TCC	GTG	AT
Y13614_PssA_	:	AAACAGT	TTGCAT	GGAAG	ACG	GGAAT	GCAT	CTGT	CCAT	GAA	GAC	TAC	AA	TCC	GTG	AT
AY258503_EpeA_	:	AAACAGT	TTGCAT	GGAAG	ACG	GGAAT	GCAT	CTGT	CCAT	GAA	GAC	TAC	AA	TCC	GTG	AT
AJ278144_EspI_	:	AAACAGT	TTGCAT	GGAAG	ACG	GGAAT	GCAT	CTGT	CCAT	GAA	GAC	TAC	AA	TCC	GTG	AT
AF056581_Pet_	:	AAACAGT	TTCTC	CTG	GAAG	ACG	GGAAT	GAAC	CTC	AC	CA	TGA	AG	GA	CTT	TA
AJ586888_Sat_	:	AAACAGT	TTCTC	CTG	GAAG	ACG	GGAAT	GAAC	CTC	AC	CA	TGA	AG	GA	CTT	TA
CP000038_SigA_	:	AAACGGT	TCGAC	TGGC	AGG	ATC	CGG	GAAT	GAG	CGT	GAC	CA	TGA	AG	GA	CTT
AF297061_EspC1_	:	AAATCTT	TTAG	CTGG	GAAG	ACCG	GGGA	ATG	GCCC	CT	GAG	CA	TGA	AG	GA	CTT
U69128_EspC2_	:	AAATCTT	TTAT	CTG	GAAG	ACCG	GGGA	ATG	GCCC	CT	GAG	CA	TGA	AG	GA	CTT
Z48219_SepA_	:	AAAGCTT	TTAG	CTGG	GAAG	ACCG	GGGA	ATG	GCT	CT	GAG	CA	TGA	AG	GA	CTT
AY604009_SepA1_	:	AAATCCT	TTAA	CTGG	GAAG	ACCG	GGGA	ATG	GCA	CT	GAT	GAA	GA	CA	TAT	TA
AY163491_EatA_	:	AAGTCAT	TTAA	CTGG	GAAG	ACCG	GGGA	ATG	CA	CT	GAT	GAA	GA	CA	TAT	TA
DD002707_Vat_	:	CAAACAT	TTAA	CTG	GAAC	GCAG	ATG	GGA	ATG	GAT	GT	CT	CA	AT	GCG	TC
AF218073_Hbp_	:	CAAACAT	TTAA	CTG	GAAC	GCAG	ATG	GGA	ATG	GAT	GT	CT	CA	AT	GCG	TC
AY876285_Boa_	:	ACAA	CGCT	GAA	ATG	GAA	TG	TA	TG	GT	AT	G	TA	TG	TA	TG
NZ_AAJV01000028_Pic-like_	:	GATTGA	ACT	GGG	CAAAA	CAT	T	CAG	TGG	TAA	G	GA	CT	GG	AG	T
citrobacter_freundii_	:	GATTGA	CGT	GGG	TAA	AA	CCT	T	CAG	TGG	TAA	G	GA	CT	GG	AG
AE005674_Pic_Shigella_	:	GGTTGA	ACT	GGG	CAAG	ACCT	T	CAG	TGG	TAA	G	GA	CT	GG	AG	T
AX276281_Pic_E.coli_	:	GATTGA	ACT	GGG	CAAG	ACCT	T	CAG	TGG	TAA	G	GA	CT	GG	AG	T
X97542_EspP_	:	TGTGGA	TGT	GGG	TAA	AT	CCT	T	CT	G	TG	TAA	G	GA	CT	GG
Y13614_PssA_	:	TGTGGA	TGT	GGG	TAA	AT	CCT	T	CT	G	TG	TAA	G	GA	CT	GG
AY258503_EpeA_	:	TGTGGA	TGT	GGG	TAA	AT	CCT	T	CT	G	TG	TAA	G	GA	CT	GG
AJ278144_EspI_	:	TGTGGA	TGT	GGG	TAA	AT	CCT	T	CT	G	TG	TAA	G	GA	CT	GG
AF056581_Pet_	:	TGTGGA	TGT	GGG	TAA	AT	CCT	T	CT	G	TG	TAA	G	GA	CT	GG
AJ586888_Sat_	:	TGTGGA	TGT	GGG	TAA	AT	CCT	T	CT	G	TG	TAA	G	GA	CT	GG
CP000038_SigA_	:	TGTGGA	TGT	GGG	TAA	AT	CCT	T	CT	G	TG	TAA	G	GA	CT	GG
AF297061_EspC1_	:	TGTGGA	TGT	GGG	TAA	AT	CCT	T	CT	G	TG	TAA	G	GA	CT	GG
U69128_EspC2_	:	TGTGGA	TGT	GGG	TAA	AT	CCT	T	CT	G	TG	TAA	G	GA	CT	GG
Z48219_SepA_	:	TGTGGA	TGT	GGG	TAA	AT	CCT	T	CT	G	TG	TAA	G	GA	CT	GG
AY604009_SepA1_	:	TGTGGA	TGT	GGG	TAA	AT	CCT	T	CT	G	TG	TAA	G	GA	CT	GG
AY163491_EatA_	:	TGTGGA	TGT	GGG	TAA	AT	CCT	T	CT	G	TG	TAA	G	GA	CT	GG
DD002707_Vat_	:	CGTTGT	TTCC	GGT	TAA	AA	CCT	T	CAG	TGG	TAA	G	GA	CT	GG	AG
AF218073_Hbp_	:	CGTTGT	TTCC	GGT	TAA	AA	CCT	T	CAG	TGG	TAA	G	GA	CT	GG	AG
AY876285_Boa_	:	CGTGG	CA	TTGG	GC	AAAA	AC	GT	T	CAG	TGG	CA	AG	CT	TAC	GG

SUBSTITUTE SHEET (RULE 26)

Figure 5B

	120	*	140	*	160	*	180	*	200	*	220
BD195283	:	GGGCACGATAACAACGGTGGTATTGCCCGTGGGGCCACGCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
AR580480	:	GGGCACGATAACAACGGTGGTATTGCCCGTGGGGCCACGCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
AX702524	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGGGCCACGCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTGAGAACAGATGTTGCCGGTAT								
AE014075	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGGGCCACGCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTGAGAACAGATGTTGCCGGTAT								
AJ586888	:	GGTCACGATAACAACGGCGGTATTGTCCGTGGTCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
AJ617685	:	GGTCACGATAACAACGGCGGTATTGTCCGTGGTCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
CP000243	:	GGGCACGATAACAACGGTGGTATTGCCCGTGGGGCCACGCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTGAGAACAGAGGTTGCCGGTAT								
CP000034	:	GGACACGATAACAACGGTGGTATTGCCCGTGGGGCCACGCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTGAGAACAGAGGTTGCCGGTAT								
AF326777	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGAGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
BD184766	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGAGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
AR204161	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGAGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
AR637510	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGAGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
BA000007	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGAGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
AE005174	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGAGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
CS148067	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGAGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTGAGAACAGAGGTTGCCGGTAT								
X16664	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGGGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTGAGAACAGAGGTTGCCGGTAT								
AX702425	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGGGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
AJ586887	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGGGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
AE005674	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGAGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
AE014073	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGAGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
AF200692	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGAGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
AY857617	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGAGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
CS148243	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGAGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
AJ303141	:	GGTCACGATAACAACGGCGGTATTGCCCGTGGAGCCACGCCCGGAAAGCAGCGGCAGCTATGGCTTCGTCCGTC	TGGAGGGTGACCTGCTCAGAACAGAGGTTGCCGGTAT								
AF233272	:	GGTCACGATAACAATGGCGGTATTGCCCGTGGGGCCACGCCCGGAAAGCAGCGGCAGCTATGGATTGTCCGTC	TGGAGGGTGACCTGATGAGAACAGAGGTTGCCGGTAT								
U24429	:	GGTCACGATAACAATGGCGGTATTGCCCGTGGGGCCACGCCCGGAAAGCAGCGGCAGCTATGGATTGTCCGTC	TGGAGGGTGACCTGATGAGAACAGAGGTTGCCGGTAT								
AX370193	:	GGTCACGATAACAATGGCGGTATTGCCCGTGGGGCCACGCCCGGAAAGCAGCGGCAGCTATGGATTGTCCGTC	TGGAGGGTGACCTGATGAGAACAGAGGTTGCCGGTAT								
BD444174	:	GGTCACGATAACAATGGCGGTATTGCCCGTGGGGCCACGCCCGGAAAGCAGCGGCAGCTATGGATTGTCCGTC	TGGAGGGTGACCTGATGAGAACAGAGGTTGCCGGTAT								
U00096	:	GGTCACGATAACAATGGCGGTATTGCCCGTGGGGCCACGCCCGGAAAGCAGCGGCAGCTATGGATTGTCCGTC	TGGAGGGTGACCTGATGAGAACAGAGGTTGCCGGTAT								
AP009048	:	GGTCACGATAACAATGGCGGTATTGCCCGTGGGGCCACGCCCGGAAAGCAGCGGCAGCTATGGATTGTCCGTC	TGGAGGGTGACCTGATGAGAACAGAGGTTGCCGGTAT								
AF233271	:	GGTCACGATAACAATGGCGGTATTGCCCGTGGGGCCACGCCCGGAAAGCAGCGGCAGCTATGGATTGTCCGTC	TGGAGGGTGACCTGATGAGAACAGAGGTTGCCGGTAT								

12 / 13

Figure 5C

BD195283	:	GTCTGTGACCGCGGGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATTATGACGGTTC	CGCGCGGCACGGTCCGGGATGATGCC	*	240	260	*	280	*	300	*
AR580480	:	GTCTGTGACCGCGGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATTATGACGGTTC	CGCGCGGCACGGTCCGGGATGATGCC								
AX702524	:	GTCGTGACCGCAGGGATATATGGTGCTGCAAGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AE014075	:	GTCGTGACCGCAGGGATATATGGTGCTGCAAGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AJ586888	:	GTCACCTGACGACAGGAGTGATGGTGCTGCAGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AJ617685	:	GTCACCTGACGACAGGAGTGATGGTGCTGCAGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
CP000243	:	GTCACCTGACGACAGGAGTGATGGTGCTGCAGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
CP000034	:	GTCACCTGACGACAGGAGTGATGGTGCTGCAGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AF326777	:	GTCTCTCACGACAGGGGTGATGGTGCCGCAAGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
BD184766	:	GTCTCTGACGACAGGGGTGATGGTGCTGCAGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AR204161	:	GTCTCTGACGACAGGGGTGATGGTGCTGCAGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AR637510	:	GTCTCTGACGACAGGGGTGATGGTGCTGCAGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
BA000007	:	GTCTCTGACGACAGGGGTGATGGTGCTGCAGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AE005174	:	GTCTCTGACGACAGGGGTGATGGTGCTGCAGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
CS148067	:	GTCTCTGACGACAGGGGTGATGGTGCTGCAGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
X16664	:	GTCTCTGACGACAGGGGTGATGGTGCTGCAGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AX702425	:	GTCACCTGACGACAGGAGTGATGGTGCTGCAGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AJ586887	:	GTCACCTGACGACAGGAGTGATGGTGCTGCAGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AE005674	:	GTCGTGACCGCGGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AE014073	:	GTCGTGACCGCGGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AF200692	:	GTCGTGACCGCGGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AY857617	:	GTCGTGACCGCGGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
CS148243	:	GTCGTGACCGCGGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AJ303141	:	GTCGTGACCGCGGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AF233272	:	GTCGTGACCGCGGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
U24429	:	GTCGTGACCGCGGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AX370193	:	GTCGTGACCGCGGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
BD444174	:	GTCGTGACCGCGGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
U00096	:	GTCGTGACCGCGGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AP009048	:	GTCGTGACCGCGGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								
AF233271	:	GTCGTGACCGCGGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGTCC	CGCGGCACGGTCCGGGATGATGCC								