



(86) Date de dépôt PCT/PCT Filing Date: 2015/05/06
(87) Date publication PCT/PCT Publication Date: 2015/11/12
(45) Date de délivrance/Issue Date: 2023/03/14
(85) Entrée phase nationale/National Entry: 2016/11/04
(86) N° demande PCT/PCT Application No.: NL 2015/050313
(87) N° publication PCT/PCT Publication No.: 2015/170979
(30) Priorités/Priorities: 2014/05/06 (EP14167256.8);
2014/05/09 (EP14167778.1)

(51) Cl.Int./Int.Cl. *C12Q 1/6809* (2018.01),
C12Q 1/686 (2018.01), *C12Q 1/689* (2018.01),
C40B 30/04 (2006.01), *G16B 20/00* (2019.01),
C07H 21/04 (2006.01)
(72) Inventeurs/Inventors:
BUDDING, ANDRIES EDWARD, NL;
SAVELKOUL, PAUL HENDRIK MARIA, NL
(73) Propriétaire/Owner:
IS-DIAGNOSTICS LTD., NL
(74) Agent: GOWLING WLG (CANADA) LLP

(54) Titre : ANALYSE D'UNE POPULATION MICROBIENNE
(54) Title: MICROBIAL POPULATION ANALYSIS

(57) **Abrégé/Abstract:**

The present invention relates to a method of typing a microbiome for having a desirable or undesirable signature, comprising analyzing the composition of the population of microorganisms in said microbiome based on taxonomic variation in the DNA sequence of the microbial 16S/23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of said microorganisms, wherein the sequences of conserved DNA regions comprised in the 16S and 23S rRNA sequences flanking said ITS region in the genome of said microorganisms comprise primer binding sites for amplification of said ITS regions.

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

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
12 November 2015 (12.11.2015)

(10) International Publication Number
WO 2015/170979 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/NL2015/050313

(22) International Filing Date:
6 May 2015 (06.05.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
14167256.8 6 May 2014 (06.05.2014) EP
14167778.1 9 May 2014 (09.05.2014) EP

(71) Applicant: **IS-DIAGNOSTICS LTD.** [NL/NL];
Troubadoursborch 59, NL-3992 BE Houten (NL).

(72) Inventors: **BUDDING, Andries Edward**; c/o van der
Boechorststraat 7, NL-1081 BT Amsterdam (NL).
SAVELKOUL, Paul Hendrik Maria; c/o van der Boe-
chorststraat 7, NL-1081 BT Amsterdam (NL).

(74) Agent: **JANSEN, C.M.**; V.O., Carnegieplein 5, 2517 KJ
Den Haag (NL).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG,
MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM,
PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC,
SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments (Rule 48.2(h))

(54) Title: MICROBIAL POPULATION ANALYSIS

(57) Abstract: The present invention relates to a method of typing a microbiome for having a desirable or undesirable signature, comprising analyzing the composition of the population of microorganisms in said microbiome based on taxonomic variation in the DNA sequence of the microbial 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of said microorganisms, wherein the sequences of conserved DNA regions comprised in the 16S and 23S rRNA sequences flanking said ITS region in the genome of said microorganisms comprise primer binding sites for amplification of said ITS regions.



WO 2015/170979 A1

Title: Microbial population analysis

5 FIELD OF THE INVENTION

The invention is in the field of medical and environmental diagnostics. More specifically, the invention relates to methods of typing and/or monitoring of specific diseases and/or environmental conditions through analysis of the microbial population composition associated with a specific disease condition or environmental condition using polymorphic nucleic acid sequences specific for individual members of the microbial population as a tool for analyzing the microbial population composition. The invention further provides means and methods for calibrating analytical results of microbial population analysis, in particular results of amplification reactions involving amplification of polymorphic nucleic acid sequences, in order to accurately and reproducibly quantify the initial template concentrations in such amplification reactions. The invention further provides methods for typing or analyzing the composition of a population of microorganisms using the said calibration means, and to kits for use in any of the methods of the invention.

BACKGROUND OF THE INVENTION

The use of molecular markers in the assessment and management of disease is becoming more and more common. Some molecular signatures are used diagnostically to help pathologists classify the disease. Molecular markers generally represent biological molecules that originate from the patient's own body cells. But apart from the tissues and organs of the body itself, the microbial community in intimate contact with a patient, such as that of the gastrointestinal tract, provides enormous potential as a reservoir of diagnostically significant information.

Although the correlation between the presence of pathogenic microorganisms and the occurrence of pathogenic disease is diagnostically well accepted, in the case of non-pathogenic disease or disease of unclear pathogenesis, significant correlations with any particular microbial species are rare, and unless a causal relationship between the disease and a specific microbe is established, any correlation may be viewed as having low diagnostic yield.

Analysis of the complex microbial flora of the human intestine is challenging and has hitherto been of limited value in the diagnosis and monitoring of diseases or the effect of therapy. Even PCR-based profiling or “fingerprinting” of microbial populations has thus far aided sparsely in the diagnosis and/or monitoring of diseases. The reason is that these microbial populations differ greatly between individual patients, preventing the ready recognition of specific disease signatures. Although a prevalence of certain species within the colonic microbiome of patients may sometimes be recognized when a comparison is made with healthy subjects, the overall composition of the microbial population is considered to vary too much between individual subjects to allow the identification of disease-specific signatures of the microbiome. As long as it is impossible to assess the meaning of diversity in one population of microorganisms to another, the step from correlation to diagnosing is seriously impeded.

Cultivation techniques have proven inadequate for the purpose of qualitatively describing complex microbial populations, and PCR –based methods are considered to provide a better insight in the true presence of individual members in the community, even if these members are uncultivable. Nonetheless, PCR-based techniques have also shown to have low reproducibility. The reason for this is not entirely clear. The present inventors have noted that two independent PCR amplification reactions performed on different samples are notoriously difficult to compare when

attempting to correlate the amount of template present to the amount of products obtained.

Therefore, there is still a need for tools that can improve the reproducibility of PCR techniques applied to mixed microbial populations.

5 Hence, there is also still a need for tools that can aid in classifying disease based on microbial population composition analysis. It is an aim of the present invention to overcome the problems of the prior art microbial profiling techniques. It is further an aim of the present invention to provide microbial profiles that can be used to distinguish healthy from diseased
10 conditions.

SUMMARY OF THE INVENTION

The present inventors have now surprisingly found that disease-specific profiles of the population composition of microorganisms can be
15 obtained, despite inter-patient variability. The present inventors have further found that nucleic acid amplification-based methods for microbial population profiling can actually be used in typing or classifying of certain diseases. The present inventors have further found that embodiments of such typing methods can benefit from an internal amplification standard
20 that allows for the correction of length-specific amplification efficiencies in a single amplification reaction, thereby allowing for the removal of inter-sample variability in the amplification efficiency of sequence-length polymorphisms.

The present inventors have established new methods for nucleic
25 acid amplification-based analysis of complex microbial populations that allow for the determination of initial amounts of specific microbial nucleic acids in a sample and to thereby quantify individual members in a mixed microbial population more accurately and reproducibly. The inventors have further found that a reference database of reference profiles can be

determinative in the finding of significant correlations between the microbiome and specific diseases.

- In one aspect, the present invention provides a method for diagnosing or monitoring a disease or condition in a patient comprising
- 5 typing a microbiome from said patient for having a healthy or diseased signature, comprising analyzing the composition of the population of microorganisms in said microbiome based on taxonomic variation in the DNA sequence of the microbial 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of said microorganisms, wherein the
- 10 sequences of conserved DNA regions comprised in the 16S and 23S rRNA sequences flanking said ITS region in the genome of said microorganisms comprise primer binding sites for amplification of said ITS regions, said analyzing comprising the steps of:
- a) providing a sample of genomic DNA from the microorganisms in a
 - 15 microbiome;
 - b) performing a PCR amplification reaction on said sample of genomic DNA using at least one set of PCR amplification primers directed to said flanking conserved DNA regions to thereby amplify and provide amplification products of said ITS regions comprised in said genomic
 - 20 DNA sample;
 - c) analyzing said amplification products based on length differences in said amplification products to thereby provide a test signature of the composition of the population of microorganisms in said microbiome;
 - d) comparing said test signature with at least one reference signature of a
 - 25 microbiome of a healthy patient and/or with at least one reference signature of a microbiome of a diseased patient, preferably by clustering of ITS profiles, and classifying the test signature as a signature of a microbiome of a healthy patient or as a signature of a microbiome of a diseased patient,
 - 30 wherein said at least one set of PCR amplification primers comprises:

- a set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of microorganisms belonging to the phylum *Firmicutes*, and
- a set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of microorganisms belonging to the phylum *Bacteroidetes*.

In another aspect, the present invention provides a method of typing a microbiome for having a desirable or undesirable signature, comprising analyzing the composition of the population of microorganisms in said microbiome based on taxonomic variation in the DNA sequence of the microbial 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of said microorganisms, wherein the sequences of conserved DNA regions comprised in the 16S and 23S rRNA sequences flanking said ITS region in the genome of said microorganisms comprise primer binding sites for amplification of said ITS regions, said analyzing comprising the steps of:

- a) providing a sample of genomic DNA from the microorganisms in a microbiome;
- b) performing a PCR amplification reaction on said sample of genomic DNA using at least one set of PCR amplification primers directed to said flanking conserved DNA regions to thereby amplify and provide amplification products of said ITS regions comprised in said genomic DNA sample;
- c) analyzing said amplification products based on length differences in said amplification products to thereby provide a test signature of the composition of the population of microorganisms in said microbiome;
- d) comparing said test signature with at least one reference signature of a desirable microbiome and/or with at least one reference signature of an undesirable microbiome, preferably by clustering of ITS profiles, and

classifying the test signature as a signature of a desirable microbiome
or as a signature of an undesirable microbiome,

wherein said at least one set of PCR amplification primers comprises:

- a set of PCR amplification primers for amplifying the 16S–23S rRNA
5 internal transcribed spacer (ITS) regions in the genomic DNA of
microorganisms belonging to the phylum *Firmicutes*, and
- a set of PCR amplification primers for amplifying the 16S–23S rRNA
internal transcribed spacer (ITS) regions in the genomic DNA of
microorganisms belonging to the phylum *Bacteroidetes*.

10 In preferred embodiments of the above method, said at least one
set of PCR amplification primers further comprises:

- a set of PCR amplification primers for amplifying the 16S–23S rRNA
internal transcribed spacer (ITS) regions in the genomic DNA of
microorganisms belonging to the phylum *Proteobacteria*.

15 In preferred embodiments of the above method, said set of PCR
amplification primers for amplifying the 16S–23S rRNA internal
transcribed spacer (ITS) regions in the genomic DNA of the phyla
Firmicutes and *Bacteroidetes* comprises the use of:

- a) the forward primer 5'-CTGGATCACCTCCTTTCTAWG-3'
20 comprising a first fluorescent label,
- b) the forward primer 5'-CTGGAACACCTCCTTTCTGGA-3'
comprising a second fluorescent label;
- c) and three unlabeled reverse primers 5'-
AGGCATCCACCGTGCGCCCT-3'; 5'-AGGCATTCACCRGTGCGCCCT-3'; and
25 5'-AGGCATCCRCCATGCGCCCT-3'.

In preferred embodiments of the above method, said set of PCR
amplification primers for amplifying the 16S–23S rRNA internal
transcribed spacer (ITS) regions in the genomic DNA of the phylum
Proteobacteria comprises the use of:

- 30 a) the forward primer 5'-CCGCCCCGTCACACCATGG-3'

b) at least one of the reverse primers selected from the group consisting of
 5'-AATCTCGGTTGATTTCTTTTCCT-3',
 5'-AATCTCGGTTGATTTCTTCTCCT-3',
 5'-AATCTCTTTTGATTTCTTTTCCTCG-3',
 5 5'-AATCTCATTTGATGTCTTTTCCTCG-3',
 5'-AATCTCTTTTGATTTCTTTTCCTTCG-3',
 5'-AATCTCTCTTGATTTCTTTTCCTTCG-3',
 5'-AATCTCAATTGATTTCTTTTCCTAAGG-3',

wherein at least one of said primers comprises a fluorescent label.

10 In preferred embodiments of the above method, said microbiome is the microbiome of the respiratory tract, oral cavity, skin, gastrointestinal tract, or urogenital tract or that of urine, saliva, sputum, pus, wound fluid, or feces, preferably of the human body, or the microbiome associated with soil, waterbodies or plants.

15 In preferred embodiments of the above method, said undesirable microbiome is

a) the intestinal microbiome of a subject suffering from:

- a digestive tract or gastrointestinal disorder including inflammatory bowel disease (IBD), diverticulitis, irritable bowel syndrome (IBS), coeliac, lactose
 20 intolerance, Necrotising Enterocolitis (NEC), Clostridium Difficile Associated Diarrhea and colorectal cancer, or a comorbid disorder selected from the group consisting of attention-deficit/hyperactivity disorder (ADHD), obsessive-compulsive disorder (OCD), anxiety, stress, eating disorder, major depressive disorder (MDD), bipolar disorder, depression,
 25 and schizophrenia,
 - a systemic disease including arthritis, sarcoidosis, mixed connective tissue disease, spondylitis ankylopoetica, osteoporosis, juvenile idiopathic arthritis, osteoarthritis, rheumatoid arthritis, sepsis, Behçet's disease, Sjögren's syndrome, fibromyalgia, scleroderma, Raynaud's phenomenon, and
 30 systemic lupus erythematosus,

- a skin disorder including psoriasis, eczema, acne, or rosacea,
- b) the skin microbiome of a subject suffering from a skin disorder including psoriasis, eczema, acne, or rosacea,
- c) a wound microbiome of a subject suffering from chronic wound, including
- 5 a diabetic ulcer, or ulcer associated with vascular disease,
- d) a urogenital tract microbiome or vaginal flora of a female having low fertility or having a low probability of success of an *in vitro* fertilization or embryo transfer procedure, or suffering from bacterial vaginosis,
- e) an oral microbiome of a subject suffering from an oral, nasal or
- 10 oropharyngeal disorder including periodontitis, periimplantitis, and oro-nasopharyngeal carcinoma,
- f) a pulmonary microbiome in a pulmonary sample selected from bronchoalveolar lavage, a sputum sample and a lung biopsy of a subject suffering from a disorder that affects the upper or lower respiratory tract,
- 15 including respiratory infection, chronic obstructive pulmonary disease (COPD), asthma, cystic fibrosis and lung cancer,
- g) a microbiome in a liquor sample, a pleural sample, a blood sample, a urine sample, an abscess sample, or a tissue sample of an organ in a subject suspected of suffering from a microbial infection,
- 20 h) a microbiome in a sample of environmental, plant, animal or food origin or in a sample of a pharmaceutical or chemical product intended to be devoid of microbes or microbial DNA.

In another aspect, the present invention provides a method of analyzing the composition of a microbiome based on taxonomic variation in

25 the DNA sequence of the microbial 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the microorganisms in said microbiome, wherein the sequences of conserved DNA regions comprised in the 16S and 23S rRNA sequences flanking said ITS region in the genomic DNA of said microorganisms comprise primer binding sites for amplification

30 of said ITS regions, said method comprising the steps of:

- a) providing a sample of genomic DNA from a microbiome;
- b) providing a PCR calibrator system, comprising a set of PCR amplification primers at least one of which primers comprises a label, and a set of at least two PCR calibrators, each PCR calibrator consisting of a DNA fragment
- 5 comprising a spacer region having a DNA sequence of a given length flanked by upstream and downstream adapter DNA sequences that comprise primer binding sites for binding of said PCR amplification primers wherein said set of PCR amplification primers is for PCR amplifying the spacer region DNA sequence of all PCR calibrators in said set of at least two PCR calibrators,
- 10 wherein the spacer region DNA sequence comprised in each of said PCR calibrators in said set of at least two PCR calibrators is of a different length, and wherein each PCR calibrator in said set of at least two PCR calibrators is present in equal amount or in a known amount relative to other PCR calibrators in said set;
- 15 c) adding said set of at least two PCR calibrators from said PCR calibrator system to said sample of genomic DNA;
- d) performing a PCR amplification reaction on said sample of genomic DNA comprising said set of at least two PCR calibrators using said set of PCR amplification primers from said PCR calibrator system as a first set of
- 20 amplification primers to amplify and provide amplification products of said ITS region(s) comprised in said set of at least two PCR calibrators, and using at least a second set of PCR amplification primers directed to said flanking conserved DNA regions to thereby co-amplify and provide amplification products of said ITS regions comprised in said sample of
- 25 genomic DNA, and;
- e) providing a standard curve by determining the PCR amplification efficiency of each of said at least two PCR calibrators from said PCR calibrator system in said PCR amplification reaction of step d) and expressing said PCR amplification efficiency as a function of the length of
- 30 the DNA sequence of the ITS region;

- f) determining the length-specific amplification efficiency for ITS regions of different length comprised in said genomic DNA sample and amplified in step d) using the standard curve as provided in step e);
- g) determining the abundance of microbial 16S–23S rRNA internal transcribed spacer (ITS) regions of different length in said microbiome using the length-specific amplification efficiencies determined in step f), and
- h) analyzing the composition of a population of microorganisms based on the abundances of ITS regions of different length determined in step g).

In preferred embodiments of the above method, said standard curve is based on at least five PCR calibrators of different length ranging in length from 50 to 1200 bps.

In preferred embodiments of the above method, said step d) of performing a PCR amplification reaction on said sample of genomic DNA using at least a set of PCR amplification primers directed to said flanking conserved DNA regions comprises the use of a set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phyla *Firmicutes* and *Bacteroidetes* and a set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phylum *Proteobacteria*.

In preferred embodiments of the above method, said set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phyla *Firmicutes* and *Bacteroidetes* comprises the use of:

- a) the forward primer 5'-CTGGATCACCTCCTTTCTAWG-3' comprising a first fluorescent label,
- b) the forward primer 5'-CTGGAACACCTCCTTTCTGGA-3' comprising a second fluorescent label;

c) and three unlabeled reverse primers 5'-AGGCATCCACCGTGCGCCCT-3'; 5'-AGGCATTCACCRGTGCGCCCT-3'; and 5'-AGGCATCCRCCATGCGCCCT-3'.

In preferred embodiments of the above method, said set of PCR
 5 amplification primers for amplifying the 16S–23S rRNA internal
 transcribed spacer (ITS) regions in the genomic DNA of the phylum
Proteobacteria comprises the use of:
 a) the forward primer 5'-CCGCCCCGTCACACCATGG-3'
 b) at least one of the reverse primers selected from the group consisting of
 10 5'-AATCTCGGTTGATTTCTTTTCCT-3',
 5'-AATCTCGGTTGATTTCTTCTCCT-3',
 5'-AATCTCTTTTGATTTCTTTTCCTCG-3',
 5'-AATCTCATTTGATGTCTTTTCCTCG-3',
 5'-AATCTCTTTTGATTTCTTTTCCTTCG-3',
 15 5'-AATCTCTCTTGATTTCTTTTCCTTCG-3',
 5'-AATCTCAATTGATTTCTTTTCCTAAGG-3',
 wherein at least one of said primers comprises a fluorescent label.

In other preferred embodiments of the above method, said
 microbiome is the microbiome of the respiratory tract, oral cavity, skin,
 20 gastrointestinal tract, or urogenital tract or that of urine, saliva, sputum,
 pus, wound fluid, or feces, preferably of the human body, or the microbiome
 associated with soil, waterbodies or plants.

In another aspect, the present invention provides a method for
 diagnosing or monitoring a digestive tract or gastrointestinal disorder in a
 25 patient, wherein the disorder is selected from the group consisting of
 digestive tract or gastrointestinal disorders associated with inflammatory
 bowel disease (IBD), diverticulitis, irritable bowel syndrome (IBS), coeliac,
 lactose intolerance, Necrotising Enterocolitis (NEC), *Clostridium Difficile*
 Associated Diarrhea and colorectal cancer, or comorbid disorder selected
 30 from the group consisting of attention-deficit/hyperactivity disorder

(ADHD), obsessive-compulsive disorder (OCD), anxiety, stress, eating disorder, major depressive disorder (MDD), bipolar disorder, depression, and schizophrenia, said method comprising performing the method of the present invention on a sample of the intestinal flora of a patient, and
5 optionally treating the patient by pharmacological intervention.

In another aspect, the present invention provides method for diagnosing or monitoring a systemic disease in a patient selected from the group consisting of arthritis, sarcoidosis, mixed connective tissue disease, spondylitis ankylopoetica, osteoporosis, juvenile idiopathic arthritis,
10 osteoarthritis, rheumatoid arthritis, sepsis, Behçet's disease, Sjögren's syndrome, fibromyalgia, sclerodermia, Raynaud's phenomenon, and systemic lupus erythematosus, said method comprising performing the method of any one of claims 1-13 on a sample of the intestinal flora of a patient, and optionally treating the patient by pharmacological intervention.

15 In another aspect, the present invention provides a method for diagnosing or monitoring a skin disorder or chronic wound in a patient, said method comprising performing the method of the present invention on a sample of the skin, wound flora or intestinal flora of a patient, and optionally treating the patient by pharmacological intervention.

20 In another aspect, the present invention provides a method of determining fertility or the probability of success of an *in vitro* fertilization or embryo transfer procedure in a female, said method comprising performing the method of the present invention on a sample of the vaginal flora, and optionally treating the patient by pharmacological intervention.

25 In another aspect, the present invention provides a method for diagnosing or monitoring bacterial vaginosis, said method comprising performing the method of the present invention on a sample of the vaginal flora, and optionally treating the patient by pharmacological intervention, preferably wherein the pharmacological intervention comprises the systemic

or topical administration of an antibiotic, preferably metronidazole, clindamycin, amoxicillin-clavulanate or fluconazole.

In another aspect, the present invention provides a method for diagnosing or monitoring an oral, nasal or oropharyngeal disorder selected from periodontitis, periimplantitis, and oro-nasopharyngeal carcinoma said
5 method comprising performing the method of the present invention on a sample of the oral flora, and optionally treating the patient by pharmacological intervention, preferably wherein the pharmacological intervention comprises the systemic, topical or local administration of an
10 antibiotic or antineoplastic agent.

In another aspect, the present invention provides a method for diagnosing or monitoring a disorder that affects the upper or lower respiratory tract, selected from respiratory infection, chronic obstructive pulmonary disease (COPD), asthma, cystic fibrosis and lung cancer, said
15 method comprising performing the method of the present invention on a pulmonary sample selected from bronchoalveolar lavage, a sputum sample and a lung biopsy, and optionally treating the patient by pharmacological intervention, preferably wherein the pharmacological intervention comprises the systemic, topical or local administration of an antibiotic
20 agent, anti-inflammatory agent or antineoplastic agent.

In another aspect, the present invention provides a method for diagnosing or monitoring a microbial infection, said method comprising performing the method of the present invention on a liquor sample, a pleural sample, a blood sample, a urine sample, an abscess sample, or a
25 tissue sample of an organ, and optionally treating the patient by pharmacological intervention, preferably wherein the pharmacological intervention comprises the systemic, topical or local administration of an antibiotic agent.

In another aspect, the present invention provides a method for
30 determining the microbial status of a sample, wherein said sample is of

environmental, plant, animal or food origin or a sample of a pharmaceutical or chemical product intended to be devoid of microbes or microbial DNA, said method comprising performing the method of the present invention on a sample from said source, and optionally treating the source from which
5 the sample was obtained, wherein said status is indicative of sterility of said sample, quality of the environment, microbial safety of food, or the health of plant or animal, preferably wherein the treatment comprises the reduction of microbial growth in said sample source or elimination of microbes or microbial DNA from said sample source.

10 In preferred embodiments of all of the above aspects and embodiments, the test signature is characterized in having an increased or decreased diversity in the phylum *Firmicutes*, *Bacteroidetes* and/or *Proteobacteria* compared to the reference signature as calculated using the Shannon index.

15 In another aspect, the present invention provides a PCR calibrator system, comprising a set of PCR amplification primers at least one of which primers comprises a label, and a set of at least two PCR calibrators, each PCR calibrator consisting of a DNA fragment comprising a spacer region having a DNA sequence of a given length flanked by upstream
20 and downstream adapter DNA sequences that comprise primer binding sites for binding of said PCR amplification primers wherein said set of PCR amplification primers is for PCR amplifying the spacer region DNA sequence of all PCR calibrators in said set of at least two PCR calibrators, wherein the spacer region DNA sequence comprised in each of said PCR
25 calibrators in said set of at least two PCR calibrators is of a different length, and wherein each PCR calibrator in said set of at least two PCR calibrators is present in equal amount or in a known amount relative to other PCR calibrators in said set.

In a preferred embodiment of said aspect, the said spacer region
30 DNA sequence is the sequence of at least a part of a microbial 16S-23S

rRNA internal transcribed spacer (ITS) region, and wherein said adapter DNA sequences in said set of at least two PCR calibrators replace the DNA sequences of the conserved DNA regions comprised in the 16S and 23S rRNA gene sequences upstream and downstream of said ITS region in the microbial genome from which said ITS region originates and wherein said adapter DNA sequences have less than 30% sequence identity with the DNA sequences of said conserved DNA regions.

In another preferred embodiment of said aspect, said spacer region DNA sequence is an artificial or semi-artificial DNA sequence of a given length, wherein each spacer region flanked by upstream and downstream adapter DNA sequences in said set of at least two calibrators has a different length. In a highly preferred embodiment, said spacer region DNA sequence consists of the sequence of a part of a microbial 16S–23S rRNA internal transcribed spacer (ITS) region, more preferably, it is at least a part of the consensus DNA sequence of at least 2, 3, 4, 5, or more microorganisms present in the microbiome that is to be analysed.

In another aspect, the present invention provides a set of at least two PCR calibrators as defined herein above, wherein said set is comprised in a single replicon or single amplifiable template, preferably a DNA plasmid. Such a plasmid, when replicated, allows the provision of said at least two PCR calibrators in exact equal amounts.

In another aspect, the present invention provides a method of analyzing the composition of a microbiome based on taxonomic variation in the DNA sequence of the microbial 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the microorganisms in said microbiome, wherein the sequences of conserved DNA regions comprised in the 16S and 23S rRNA sequences flanking said ITS region in the genomic DNA of said microorganisms comprise primer binding sites for amplification of said ITS regions, said method comprising the steps of:

a) providing a sample of genomic DNA from a microbiome;

- b) providing a PCR calibrator system as defined herein above;
- c) adding said set of at least two PCR calibrators from said PCR calibrator system to said sample of genomic DNA;
- d) performing a PCR amplification reaction on said sample of
5 genomic DNA comprising said set of at least two PCR calibrators using said set of PCR amplification primers from said PCR calibrator system as a first set of amplification primers to amplify and provide amplification products of said ITS region(s) comprised in said set of at least two PCR calibrators, and using at least a second set of PCR amplification primers directed to said
10 flanking conserved DNA regions to thereby co-amplify and provide amplification products of said ITS regions comprised in said sample of genomic DNA, and;
- e) providing a standard curve by determining the PCR amplification efficiency of each of said at least two PCR calibrators from
15 said PCR calibrator system in said PCR amplification reaction of step d) and expressing said PCR amplification efficiency as a function of the length of the DNA sequence of the ITS region;
- f) determining the length-specific amplification efficiency for ITS regions of different length comprised in said genomic DNA sample and
20 amplified in step d) using the standard curve as provided in step e);
- g) determining the abundance of microbial 16S–23S rRNA internal transcribed spacer (ITS) regions of different length in said microbiome using the length-specific amplification efficiencies determined in step f), and
25 h) analyzing the composition of a population of microorganisms based on the abundances of ITS regions of different length determined in step g).

In a preferred embodiment of a method of analyzing the composition of a microbiome as described above, said standard curve is

based on at least five PCR calibrators of different length, preferably said PCR calibrators ranging in length from 50 to 1200 bps.

In another preferred embodiment of a method of analyzing the composition of a microbiome as described above, said step d) of performing a PCR amplification reaction on said sample of genomic DNA using at least a second set of PCR amplification primers directed to said flanking conserved DNA regions comprises the use of a set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phyla *Firmicutes* and *Bacteroidetes* and a set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phylum *Proteobacteria*. Preferably, said set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phyla *Firmicutes* and *Bacteroidetes* comprises the use of:

- a) the forward primer 5'-CTGGATCACCTCCTTTCTAWG-3' comprising a first fluorescent label,
- b) the forward primer 5'-CTGGAACACCTCCTTTCTGGA-3' comprising a second fluorescent label;
- c) and three unlabeled reverse primers 5'-AGGCATCCACCGTGCGCCCT-3'; 5'-AGGCATTCACCRGTGCGCCCT-3'; and 5'-AGGCATCCRCCATGCGCCCT-3'. Preferably, said set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phylum *Proteobacteria* comprises the use of
 - a) the forward primer 5'- CCGCCCGTCACACCATGG-3' comprising a first fluorescent label,
 - b) and seven unlabeled reverse primers
 - 5'-AATCTCGGTTGATTTCTTTTCCT-3';
 - 5'-AATCTCGGTTGATTTCTTCTCCT-3';

5'-AATCTCTTTTGATTTCTTTTCCTCG-3';
 5'-AATCTCATTTGATGTCTTTTCCTCG-3';
 5'-AATCTCTTTTGATTTCTTTTCCTTCG-3';
 5'-AATCTCTCTTGATTTCTTTTCCTTCG-3'; and
 5'-AATCTCAATTGATTTCTTTTCCTAAGG-3'

In yet another preferred embodiment of a method of analyzing the composition of a microbiome as described above, said microbiome is the microbiome of the respiratory tract, oral cavity, skin, gastrointestinal tract, or urogenital tract, or that of urine, saliva, sputum, pus, wound fluid, or feces, preferably of the human body, or the microbiome associated with soil, waterbodies or plants.

In another aspect, the present invention provides a method for diagnosing or monitoring a digestive tract or gastrointestinal disorder in a patient, wherein the disorder is selected from the group consisting of digestive tract or gastrointestinal disorders associated with inflammatory bowel disease (IBD), diverticulitis, irritable bowel syndrome (IBS), coeliac, lactose intolerance, Necrotising Enterocolitis (NEC), Clostridium Difficile Associated Diarrhea and colorectal cancer, or comorbid disorder selected from the group consisting of attention-deficit/hyperactivity disorder (ADHD), obsessive-compulsive disorder (OCD), anxiety, stress, eating disorder, major depressive disorder (MDD), bipolar disorder, depression, and schizophrenia, said method comprising performing the method of analyzing the composition of a microbiome as described above using the PCR calibrator system on a sample of the intestinal flora, and optionally treating the patient by pharmacological intervention.

The term "monitoring" as used herein refers to supervising or controlling the disease state of a subject or a microbial state of an environment. The term includes reference detection of any changes in the disease state of a subject or a microbial state of an environment. Preferably, the monitoring methods in aspects of this invention are applied for

monitoring patients in which a change in their disease state is not evident (i.e. the patient does not show obvious signs of a change in the disease state, or the patient is suspected of having a certain disease). The term “monitoring” in this aspect, and in other aspects of this invention, refers to
5 the detection of any changes over time, for instance as part of longitudinal follow up studies to detect changes in the microbiome resulting from therapy or intervention.

In a preferred embodiment of the method as described above, said sample of intestinal flora is a sample obtained from small or large intestine,
10 preferably colon or rectum, more preferably obtained in the form of a feces sample or rectal swab or in the form of a biopsy specimen of gastrointestinal mucosa.

In a further preferred embodiment of a method as described above, IBD is selected from Crohn’s disease, colitis ulcerosa, collagenous
15 colitis, lymphocytic colitis, ischemic colitis, microscopic colitis, Behçet’s syndrome, diversion colitis, diverticular colitis, eosinophilic colitis, and radiation colitis.

In yet another preferred embodiment of a method for diagnosing or monitoring a digestive tract or gastrointestinal disorder the digestive
20 tract or gastrointestinal disorder is treated by pharmacological intervention with therapeutic agents selected from an anti-inflammatory agent and an antibiotic agent, TNF-alfa inhibitors, including infliximab and adalimumab, and 5-ASA formulations, including sulfasalazine and mesalazine.

In another aspect, the present invention provides a method for
25 diagnosing or monitoring a systemic disease in a patient, wherein the systemic disease is selected from the group consisting of arthritis, sarcoidosis, mixed connective tissue disease, spondylitis ankylopoetica, osteoporosis, juvenile idiopathic arthritis, osteoarthritis, rheumatoid arthritis, sepsis, Behçet’s disease, Sjögren’s syndrome, fibromyalgia,
30 sclerodermia, Raynaud’s phenomenon, and systemic lupus erythematosus,

said method comprising performing the method of analyzing the composition of a microbiome as described above using the PCR calibrator system on a sample of said patient, preferably on a sample of the intestinal flora, and optionally treating the patient by pharmacological intervention.

5 In a preferred embodiment of the method as described above, said sample of intestinal flora is a sample obtained from small or large intestine, preferably colon or rectum, more preferably obtained in the form of a feces sample or rectal swab or in the form of a biopsy specimen of gastrointestinal mucosa.

10 In yet another preferred embodiment of a method for diagnosing or monitoring a digestive tract or gastrointestinal disorder the digestive tract or gastrointestinal disorder is treated by pharmacological intervention with therapeutic agents selected from an anti-inflammatory agent and an antibiotic agent.

15 In yet another preferred embodiment of a method for diagnosing or monitoring a digestive tract or gastrointestinal disorder as described above, the comorbid disorder is attention-deficit/hyperactivity disorder (ADHD) and the pharmacological intervention comprises treating the patient with a methylphenidate stimulant, preferably selected from the
20 group consisting of Adderall, Concerta, Vyvanse, and Ritalin.

 In yet another preferred embodiment of a method for diagnosing or monitoring a digestive tract or gastrointestinal disorder as described above, the comorbid disorder is selected from obsessive-compulsive disorder (OCD), anxiety, stress, major depressive disorder (MDD), bipolar disorder
25 symptoms, depression, and schizophrenia, and the pharmacological intervention comprises treating the patient with an antidepressant, preferably selective serotonin reuptake inhibitors (SSRIs), more preferably fluoxetine, citalopram, escitalopram, paroxetine, or sertraline; or an antipsychotic selected from the group consisting of aripiprazole, asenapine,

clozapine, iloperidone, olanzapine, paliperidone, quetiapine, risperidone, and ziprasidone.

In yet another preferred embodiment of a method for diagnosing or monitoring a digestive tract or gastrointestinal disorder as described
5 above, the eating disorder is anorexia nervosa or bulimia nervosa, and the pharmacological intervention comprises the administration of an antidepressant, preferably selective serotonin reuptake inhibitors (SSRIs), more preferably fluoxetine, citalopram, escitalopram, paroxetine, or sertraline; an antihistamine; or an antipsychotic, preferably selected from
10 the group consisting of aripiprazole, asenapine, clozapine, iloperidone, olanzapine, paliperidone, quetiapine, risperidone, and ziprasidone.

In another aspect, the present invention provides a method for diagnosing or monitoring a skin disorder or chronic wound in a patient, said method comprising performing the method of analyzing the composition of a
15 microbiome as described above using the PCR calibrator system on a sample of the skin or wound flora, and optionally treating the patient by pharmacological intervention.

In a preferred embodiment of the method for diagnosing or monitoring a skin disorder or chronic wound in a patient, said skin disorder
20 is psoriasis, eczema, acne, rosacea, diabetic ulcer, and ulcers associated with vascular disease and wherein the pharmacological intervention comprises the administration of an antibiotic agent or anti-inflammatory agent.

Antibiotic agents in aspects of this invention are preferably selected from the group consisting of amikacin, amoxicillin, anidulafungin,
25 azithromycin, bacitracin, benzylpenicillin, cephalixin, cefazolin, cefotaxime, ceftaroline, ceftazidime, ceftriaxone, cefuroxime, ciprofloxacin, clarithromycin, clindamycin, clotrimazole, co-amoxiclav, dalfopristin, daptomycin, doxycycline, ertapenem, erythromycin, ethambutol, flucloracillin, fluconazole, fosfomycin, framycetin, fusidic acid, gentamycin,
30 imipenem, isoniazid, itraconazole, ketoconazole, metronidazole, meropenem,

miconazole, moxifloxacin, mupirocin, neomycin sulphate, nitrofurantoin, ofloxacin, phenoxymethylpenicillin, piperacillin, piperacillin/tazobactam, pyrazinamide, quinupristin, rifampicin, rifaximin, teicoplanin, tetracycline, tobramycin, tigecycline, trimethoprim, sulphamethoxazole, 5 vancomycin and combinations thereof.

Anti-inflammatory agents in aspects of this invention are preferably selected from the group consisting of cortisone, hydrocortisone, prednisone and prednisolone, methylprednisolone, dexamethasone, triamcinolone, mometasone, fluticasone, betamethasone, halometasone, 10 desonide and combinations thereof.

In another aspect, the present invention provides a method of determining fertility or the probability of success of an *in vitro* fertilization or embryo transfer procedure in a female, said method comprising performing the method of analyzing the composition of a microbiome as 15 described above using the PCR calibrator system on a sample of the vaginal flora, and optionally treating the patient by pharmacological intervention.

In another aspect, the present invention provides a method for diagnosing or monitoring bacterial vaginosis, said method comprising performing the method of analyzing the composition of a microbiome as 20 described above using the PCR calibrator system of the present invention on a sample of the vaginal flora, and optionally treating the patient by pharmacological intervention, preferably the pharmacological intervention comprises the systemic or topical administration of an antibiotic agent, preferably metronidazole or clindamycin.

25 In another aspect, the present invention provides a method for diagnosing or monitoring an oral, nasal or oropharyngeal disorder selected from periodontitis, periimplantitis, and oro-nasopharyngeal carcinoma, said method comprising performing the method of analyzing the composition of a microbiome as described above using the PCR calibrator system on a sample 30 of the oral flora, and optionally treating the patient by pharmacological

intervention, by removal or revision of a dental implant, or by radiotherapy, preferably wherein the pharmacological intervention comprises the systemic, topical or local administration of an antibiotic or antineoplastic agent.

5 Antineoplastic agents in aspects of this invention are preferably selected from the group consisting of busulfan, cyclophosphamide, ifosfamide, thiotepa; mechlorethamine, melphalan, lomustine, semustine, streptozocin, capecitabine, cladribine, cytarabine, floxuridine, fludarabine, fluorouracil, mercaptopurine, methotrexate, thioguanine, bleomycin sulfate, dactinomycin, daunorubicin, doxorubicin, epirubicin, idarubicin, mitomycin, 10 mitoxantrone, pentostatin, plicamycin, streptozocin, vinblastine, vincristine, vindesine, vinorelbine tartrate, estrogens, androgens, progestins, corticosteroids, aminoglutethimide, chlorotrianisene, flutamide, goserelin, leuprolide, tamoxifen, etoposide, teniposide, paclitaxel, carboplatin, 15 cisplatin, asparaginase, dacarbazine, hydroxyurea, interferons, levamisole, mitotane, procarbazine, tretinoin, and combinations thereof.

 In another aspect, the present invention provides a method for diagnosing or monitoring a disorder that affects the upper or lower respiratory tract, selected from respiratory infection, chronic obstructive 20 pulmonary disease (COPD), asthma, cystic fibrosis and lung cancer, said method comprising performing the method of analyzing the composition of a microbiome as described above using the PCR calibrator system on a pulmonary sample selected from bronchoalveolar lavage, a sputum sample and a lung biopsy, and optionally treating the patient by pharmacological 25 intervention, preferably wherein the pharmacological intervention comprises the systemic, topical or local administration of an antibiotic, anti-inflammatory or antineoplastic agent.

 In another aspect, the present invention provides a method for diagnosing or monitoring a microbial infection, said method comprising 30 performing the method of analyzing the composition of a microbiome as

described above using the PCR calibrator system on a liquor sample, a pleural sample, a blood sample, a urine sample, an abscess sample, or a tissue sample of an organ, and optionally treating the patient by pharmacological intervention, preferably wherein the pharmacological
5 intervention comprises the systemic, topical or local administration of one or more antibiotic agents.

In another aspect, the present invention provides a method for diagnosing or monitoring a microbial infection, said method comprising performing the method of analyzing the composition of a microbiome as
10 described above using the PCR calibrator system on a liquor sample, a pleural sample, a blood sample, a urine sample, an abscess sample, or a tissue sample of an organ, and optionally treating the patient by pharmacological intervention, preferably wherein the pharmacological intervention comprises the systemic, topical or local administration of one or
15 more antibiotic agents.

In another aspect, the present invention provides a method for determining the microbial status of a sample, wherein said sample is of environmental, plant, animal or food origin, or a sample of a pharmaceutical or chemical product intended to be devoid of microbes or microbial DNA said
20 method comprising performing the method of analyzing the composition of a microbiome as described above using the PCR calibrator system on a sample from said source, and optionally treating the source from which the sample was obtained, wherein said status is indicative of sterility of said sample, quality of the environment, microbial safety of food, or the health of plant or
25 animal, preferably wherein the treatment comprises the reduction of microbial growth in said sample source.

In another aspect, the present invention provides a method of typing the intestinal flora of a subject for having a diverticulitis signature, comprising analyzing the composition of the population of microorganisms
30 in said intestinal flora based on taxonomic variation in the DNA sequence of

the microbial 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of said microorganisms, wherein the sequences of conserved DNA regions comprised in the 16S and 23S rRNA sequences flanking said ITS region in the microbial genome comprise primer binding sites for amplification of said ITS regions, said analysis comprising the steps of:

- a) providing a sample of genomic DNA from the population of microorganisms constituting the intestinal flora of a subject;
- b) performing a PCR amplification reaction on said sample of genomic DNA using at least one set of PCR amplification primers directed to said flanking conserved DNA regions to thereby amplify and provide amplification products of said ITS regions comprised in said genomic DNA sample;
- c) analyzing said amplification products to thereby provide a test signature of the composition of the population of microorganisms in said intestinal flora;
- d) comparing said test signature with a reference signature of a healthy subject and/or a subject suffering from diverticulitis and classifying the test signature as a signature of a healthy subject or as a signature of a subject suffering from diverticulitis.

In a preferred embodiment of a method of typing the intestinal flora of a subject for having a diverticulitis signature as described above, said step b) of performing a PCR amplification reaction on said sample of genomic DNA using at least one set of PCR amplification primers directed to said flanking conserved DNA regions comprises the use of a set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phyla *Firmicutes* and *Bacteroidetes* and a set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phylum *Proteobacteria*.

In another preferred embodiment of a method of typing the intestinal flora of a subject for having a diverticulitis signature as described above, said set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phyla *Firmicutes* and *Bacteroidetes* comprises the use of:

- a) the forward primer 5'-CTGGATCACCTCCTTTCTAWG-3' comprising a first fluorescent label,
- b) the forward primer 5'-CTGGAACACCTCCTTTCTGGA-3' comprising a second fluorescent label;
- 10 c) and three unlabeled reverse primers 5'-AGGCATCCACCGTGCGCCCT-3'; 5'-AGGCATTACACRTGCGCCCT-3'; and 5'-AGGCATCCRCCATGCGCCCT-3'.

In yet another preferred embodiment of a method of typing the intestinal flora of a subject for having a diverticulitis signature as described above, said set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phylum *Proteobacteria* comprises the use of:

- a) the forward primer 5'-CCGCCCCGTCACACCATGG-3'
- b) at least one of the reverse primers selected from the group consisting of
- 20 5'-AATCTCGGTTGATTTCTTTTCCT-3',
- 5'-AATCTCGGTTGATTTCTTCTCCT-3',
- 5'-AATCTCTTTTGATTTCTTTTCCTCG-3',
- 5'-AATCTCATTTGATGTCTTTTCCTCG-3',
- 5'-AATCTCTTTTGATTTCTTTTCCTTCG-3',
- 25 5'-AATCTCTCTTGATTTCTTTTCCTTCG-3',
- 5'-AATCTCAATTGATTTCTTTTCCTAAGG-3',

wherein at least one of said primers comprises a fluorescent label, preferably a fluorescent label that is different from other taxon-specific fluorescent labels in the reaction mixture.

In another preferred embodiment of a method of typing the intestinal flora of a subject for having a diverticulitis signature as described above, said method comprises the use of a PCR calibrator system as defined herein above, using the method of analyzing the composition of a microbiome as described above using the PCR calibrator system.

In another preferred embodiment of a method of typing the intestinal flora of a subject for having a diverticulitis signature as described above, the signature of a subject suffering from diverticulitis is characterized in having an increased diversity in the phylum *Proteobacteria* as calculated using the Shannon index.

In another preferred embodiment of a method of typing the intestinal flora of a subject for having a diverticulitis signature as described above, the increased diversity in the phylum *Proteobacteria* is constituted by an increased diversity in the family *Enterobacteriaceae*.

In another preferred embodiment of a method of typing the intestinal flora of a subject for having a diverticulitis signature as described above, an increased diversity in the phylum *Proteobacteria* is indicated by an increase in the presence of at least one of the species selected from the group consisting of *Escherichia coli*, *Klebsiella pneumoniae*, *Enterobacter aerogenes*, *Serratia marcescens*, *Klebsiella variicola*, *Providencia stuartii*, *Desulfovibrio* spp., *Stenotrophomonas* spp. (*Xanthomonas* spp.), *Pseudomonas aeruginosa*, *Burkholderia* spp. and *Aggregatibacter actinomycetemcomitans*.

In another preferred embodiment of a method of typing the intestinal flora of a subject for having a diverticulitis signature as described above, said sample of intestinal flora is a sample obtained from the small or large intestine, preferably colon or rectum, more preferably obtained in the form of a fecal sample or rectal swab or in the form of a biopsy specimen of gastrointestinal mucosa.

DESCRIPTION OF THE DRAWINGS

Figure 1 shows PCR calibrator profiles and potential interpretations for correction of nonlinear differences in amplification efficiency of fragments of different length as exemplified in Example 1.

- 5 Figure 1A: Typical digital profile of the five PCR calibrator fragments after amplification and length sorting by capillary gel electrophoresis. Figure 1B. Multi-linear model of amplification efficiency profile. Figure 1C. Non-linear model of amplification efficiency. Figure 1D. Example of the application of a corrective factor as estimated by the amplification function on all peaks in a
10 profile.

Figure 2 shows the principle of up-scaling of a partially inhibited sample using the PCR calibrator system as exemplified in Example 1.

- Figure 3 shows the use of the PCR calibrator system in error checking of automated length marker detection as exemplified in Example
15 1. Figure 3A: Raw capillary gel electrophoresis data for the length marker channel. Figure 3B: Length marker peaks are identified and assigned their respective length by an automated pre-processing software system. Figure 3C: Raw capillary gel electrophoresis data for the length markers including the PCR calibrator channel.

- 20 Figure 4 shows a heat map of all profiles sorted and colored by phylum for the experiment as described in Example 2.

- Figure 5 shows boxplot comparisons of within-sample diversity as calculated by Shannon index (all phyla combined and per phylum) for diverticulitis patients and control subjects for the experiment as described
25 in Example 2.

- Figure 6 shows a Principle Coordinate Analysis (PCoA) scatterplot to express between-sample diversity and displays clustering of diverticulitis patients separate from control subjects for the phylum *Proteobacteria* as explained in Example 2. Three wrongly classified samples
30 are encircled.

Figure 7 shows the partial least square-discriminant analysis (PLS-DA) scores plot for the phylum *Proteobacteria* as explained in Example 2.

Figure 8 shows ROC curves summarizing the predictive power of the PLS-DA model as described in Example 2 for clinical status per phylum (A) and for all phyla combined (B).

Figure 9 shows schematically how two subject groups were sampled in Example 3. Group A consisted of 38 subjects who underwent an elective colonoscopy. This group collected feces at home 2 to 5 days prior to the procedure. Rectal swabs and biopsies were taken during the procedure. Two swabs were stored for 2 hours at room temperature and at -20°C afterwards, one swab was immediately snap frozen. The mucosal biopsy was washed in PBS and snap frozen. Group B consisted of 10 patients with inflammatory bowel disease. This group collected feces and one rectal swab at home. Feces was stored in a sterile container and the rectal swab in RTF buffer, both at -20°C. A second rectal swab was obtained on the day of presentation at the outpatient clinic and stored in the same fashion as the first swab.

Figure 10 shows a sample IS-profile in the context of Example 3. The x-axis represents IS fragment length, the y-axis represents relative abundance of fragments in relative fluorescent units (RFU). Colors of fragments correspond to bacterial phyla. Blue peaks represent *Firmicutes/Actinobacteria*, red peaks represent *Bacteroidetes* and yellow peaks represent *Proteobacteria*. Each peak may be regarded as an operational taxonomic unit (OTU).

Figure 11 shows various comparisons of microbiota profiles expressed as R squared in the context of Example 3. All comparisons were done separately for *Firmicutes/Actinobacteria* (left) and *Bacteroidetes* (right). Figures show comparisons of all profiles. Red dots represent comparisons of samples of the same subject (intra-subject correlation).

Yellow box plots are based on all correlations, red box plot on intra-subject correlations only. A: Duplicate swab profiles stored in RTF buffer. B: Swab stored in RTF buffer vs. snap frozen swabs. C: Swabs stored in buffer vs. mucosal biopsies. D: Swabs stored in buffer vs. fecal samples. E: fecal samples vs. mucosal biopsies.

Figure 12 shows comparisons of microbiota profiles in the IBD polyclinic group expressed as R squared in the context of Example 3. All comparisons have been done separately for *Firmicutes/Actinobacteria* (left) and *Bacteroidetes* (right). Figures show comparisons of all profiles. Red dots represent comparisons of samples of the same subject (intra-subject correlation). Yellow box plots are based on all correlations, red box plots on intra-subject correlations only. A: Swabs taken by patients at home vs. swabs taken by the physician at the polyclinic. B: Swabs vs. fecal samples

Figure 13 shows the presence of the seven paropathogens detected by real-time PCR in healthy and diseased subject as described in Example 4. Paropathogens are displayed by patient category (healthy or diseased). Bar heights correspond to the number of subjects positive for that pathogen. It can be seen that five of the seven paropathogens are detected only in diseased subjects.

Figure 14 shows a Clustered Heat map depicting load as measured by quantitative real-time PCR of each paropathogen per patient as described in Example 4. Blue is lowest load, red is highest load. Columns are clustered by Euclidian distance, the dendrogram corresponds to these distances.

Figure 15 shows a Clustered Heat map depicting IS-profiles of the three major phyletic groups covered in IS-pro as described in Example 4. Note that *Firmicutes* group also includes *Actinobacteria*, *Fusobacteria* and *Verrucomicrobia*. It can be seen that there is a clear bipartite separation of microbiota profiles into the healthy and diseased groups, while there is also significant variation within these groups. It can be seen that especially more

Proteobacteria species are present in the parodontitis group, corresponding to a higher diversity within this phylum.

Figure 16 shows the Shannon diversity per phylum and for all phyla combined as described in Example 4. *Proteobacteria* and *Bacteroidetes* display a significantly higher diversity in diseased than in healthy individuals. This also results in a higher diversity for all phyla combined.

Figure 17 shows the sample separation by PLS-DA as described in Example 4. It can be clearly seen here that diseased and control samples fall into distinct groups.

Figure 18 shows the IS-Pro profile (profile view) of fecal samples of Crohn's Disease patients and control subjects as described in Example 5.

Figure 19 shows the IS-Pro profile (band view, or heat map) of fecal samples of Crohn's Disease patients and control subjects as described in Example 5.

Figure 20 shows the IS-Pro profile (profile view) of fecal samples of Ulcerative colitis patients and control subjects as described in Example 6.

Figure 21 shows the IS-Pro profile (band view) of fecal samples of Ulcerative colitis patients and control subjects as described in Example 6

Figure 22 shows the IS-Pro profile (profile view) of fecal samples of Eczema patients and control subjects as described in Example 7.

Figure 23 shows the IS-Pro profile (band view) of fecal samples of Eczema patients and control subjects as described in Example 7.

Figure 24 shows the IS-Pro profile (profile view) of fecal samples of Colon carcinoma patients and adenoma control subjects as described in Example 8.

Figure 25 shows the IS-Pro profile (*Firmicutes* band view) of fecal samples of Colon carcinoma patients and adenoma control subjects as described in Example 8.

Figure 26 shows the IS-Pro profile (profile view) of paperpoint samples of Peri-implantitis patients and control subjects as described in Example 9.

5 Figure 27 shows the IS-Pro profile (band view) of paperpoint samples of Peri-implantitis patients and control subjects as described in Example 9.

Figure 28 shows the IS-Pro profile (profile view) of a sample of a patient suffering from an Abscess as described in Example 10. Panel A shows a petri dish following incubation with the Abscess sample, exemplifying the culture result wherein no bacterial growth was observed. Panel B shows the IS profile of the Abscess sample, indicating the presence of a single bacterial species of the phylum *Firmicutes*, *Streptococcus cristatus*.

15 Figure 29 shows the IS-Pro profile (profile view) of fecal samples of Anorexia nervosa patients and control subjects as described in Example 11.

Figure 30 shows the IS-Pro profile (band view) of fecal samples of Anorexia nervosa patients and control subjects as described in Example 11.

20 Figure 31 shows the IS-Pro profile (profile view) of fecal samples of Asthma patients and control subjects as described in Example 12.

Figure 32 shows the IS-Pro profile (band view) of fecal samples of Asthma patients and control subjects as described in Example 12.

25 Figure 33 shows the IS-Pro profile (profile view) of vaginal samples of bacterial vaginosis patients and control subjects as described in Example 13.

Figure 34 shows the IS-Pro profile (band view) of vaginal samples of bacterial vaginosis patients and control subjects as described in Example 13.

Figure 35 shows the IS-Pro profile (cumulative profile view) of fecal samples of obstipation patients and control subjects as described in Example 14.

Figure 36 shows the IS-Pro profile (individual profile view) of
5 fecal samples of obstipation patients and control subjects as described in Example 14.

Figure 37 shows the IS-Pro profile (band view) of fecal samples of obstipation patients and control subjects as described in Example 14.

Figure 38 shows the IS-Pro profile (cumulative profile view) of
10 fecal samples of Irritable Bowel Syndrome patients and control subjects as described in Example 15.

Figure 39 shows the IS-Pro profile (individual profile view) of fecal samples of Irritable Bowel Syndrome patients and control subjects as described in Example 15.

Figure 40 shows the IS-Pro profile (band view) of fecal samples of
15 Irritable Bowel Syndrome patients and control subjects as described in Example 15.

Figure 41 shows the IS-Pro profile (profile view) of a cerebro-spinal fluid sample of patient suffering from meningitis as described in
20 Example 16. Panel A shows a petri dish following incubation with the liquor sample, exemplifying the culture result wherein no bacterial growth was observed. Panel B shows the IS profile of the liquor sample, indicating the presence of a single bacterial species of the phylum *Firmicutes*, *Streptococcus mitis*.

Figure 42 shows the IS-Pro profile (individual profile view) of
25 fecal samples of Rheuma patients and control subjects as described in Example 17.

Figure 43 shows the IS-Pro profile (band view) of fecal samples of Rheuma patients and control subjects as described in Example 17.

Figure 44 relates to the diagnosis, monitoring and prediction of Necrotising Enterocolitis (NEC) in prematurely born children. Figure 44A shows the IS-Pro profile (individual profile view) of fecal samples of Necrotising Enterocolitis (NEC) patients and control subjects as described in Example 18. Figure 44B shows boxplot comparisons of within-sample diversity as calculated by Shannon index (per phylum) for Necrotising Enterocolitis (NEC) patients and control subjects for the experiment as described in Example 18. Figure 44C shows ROC curves summarizing the predictive power of a PLS-DA model for clinical status per phylum and for all phyla combined. From this figure it follows that differences between patients and control subjects in the phylum *Proteobacteria* are exploited to diagnose, monitor and predict Necrotising Enterocolitis (NEC) in prematurely born children.

Figure 45 relates to the diagnosis, monitoring and, especially, prediction of the occurrence of sepsis in neonates based on analysis of the intestinal microbiota (feces). Figure 45A shows the IS-Pro profile (individual profile view) of fecal samples of sepsis patients and control subjects as described in Example 19. Figure 45B shows boxplot comparisons of within-sample diversity as calculated by Shannon index (per phylum) for sepsis patients and control subjects for the experiment as described in Example 19. Figure 45C shows ROC curves summarizing the predictive power of a PLS-DA model for clinical status per phylum and for all phyla combined. As evidenced by the Figures differences between patients and control subjects in all tested phyla, i.e. *Bacteroidetes*, *Firmicutes* and *Proteobacteria* are exploited to diagnose, monitor and predict sepsis.

Figure 46 shows the result of diagnosis of *Clostridium difficile* infection in the intestinal tract of (from top to bottom) infected and diseased patients, asymptomatic carriers and healthy (no *C. difficile*) subjects having diarrhoea. Displayed are (from left to right) the IS-pro profiles, the Phylum distribution (red: *Bacteroidetes*; yellow: *Proteobacteria*; bleu: *Firmicutes*),

and the Diagnosis. Clearly visible is the low population diversity as indicated by a low Shannon index in diarrhoea patients suffering from infection with a pathogenic *C. difficile* strain, while asymptomatic carriers of a low pathogenic *C. difficile* strain and healthy subjects maintain

5 Shannon index diversities above 3. Hence, subject having been diagnosed with a low overall population diversity in combination with *C. difficile* infection may benefit from fecal microbiota transplantation (FMT; e.g. as described by Bakken et al 2011, Clin Gastroenterol Hepatol. 9(12): 1044–1049), rather than antibiotic treatment alone, while diarrhoea patients

10 having a high overall population diversity in combination with *C. difficile* infection may benefit from antibiotic treatment.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

15 The term “microbiome”, as used herein, refers to a population of microorganisms from a particular environment, including the environment of the body or a part of the body, as well as the population of microorganisms inhabiting soils, plants and waterbodies. The term is interchangeably used to address the population of microorganisms itself

20 (sometimes referred to as the microbiota), as well as the collective genomes of the microorganisms that reside in the particular environment.

The term “environment”, as used herein, refers to all surrounding circumstances, conditions, or influences to which a population of microorganisms is exposed. The term is intended to include reference to any

25 a subject of study, hence, including environments in a subject, such as a human subject, but particularly refers to environments such as soil, a waterbody or a plant.

The term “disease or condition in a subject” may include reference to disease or condition in an environment, by preferably refers to disease or

30 condition in a human or animal subject.

The terms “IS-pro” and “IS-profiling”, as used interchangeably herein, are used in the context of a specific method of analyzing the composition of a microbiome based on taxonomic variation in the DNA sequence of the microbial 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the microorganisms in said microbiome. The technique is described in detail in Budding et al. 2010 (FASEB J., 24(11):4556-64)

and is also described in detail herein below.

The term “intergenic spacer region”, as used herein, refers to a genomic sequence located between two genes.

The terms “16S–23S rRNA internal transcribed spacer (ITS) region” and “16S-23S intergenic spacer (IS) region”, as used herein, refer to a segment of non-functional DNA situated between structural ribosomal RNA (rRNA) genes on a common precursor transcript. In literature, this region is also synonymously called 16S–23S rDNA intergenic spacer region (IS region) (Budding et al., 2010, cited above), 16S–23S rRNA intergenic spacer region (Mora et al. 2003. Microbiology 149: 807–813), 16S–23S rRNA gene internal transcribed spacer (ITS) region (Wang et al 2008. J Clin Microbiol. 46(11): 3555–3563), 16s/23s ribosomal spacer region (Barry, T. et al. 1991. PCR Methods Appl. 1:51-56. In the genome of *Escherichia coli* CFT073 (Genbank accession AE014075.1), the ITS region separating the 16S and 23S rRNA genes in one of the *rrn* operons is indicated by nucleotides numbered 236727-237160, comprising 433 bases. In many microbial species the 16S–23S ITS region contains coding sequences for tRNA genes. Multiple rRNA operons (*rrn*) may be present within the genome of a microorganism, sometimes as many as 15, which often display intragenomic heterogeneity in ITS type. The spacer regions between the 16S and 23S genes in the prokaryotic rRNA genetic loci show a significant level of length and sequence polymorphism across both genus and species lines. Pairs of priming sequences can be selected for the amplification of these

polymorphic regions from highly conserved sequences in the 16S and 23S genes occurring adjacent to these polymorphic regions.

The term “16S rRNA gene”, as used herein, refers to a DNA sequence or sequences encoding the 16S rRNA molecule.

5 The term “23S rRNA gene”, as used herein, refers to a DNA sequence or sequences encoding the 23S rRNA molecule.

The term “polymorphic DNA target region”, as used herein, refers to a DNA region varying in length and/or sequence in different taxonomic groups of microorganisms and that serves as a target for PCR amplification.

10 The term “conserved region”, as used herein, refers to a segment of nucleotide sequence of a gene or amino acid sequence of a protein that is significantly similar between various different nucleotide sequences of a gene. This term is interchangeably used with the term “conserved
15 region that (i) comprises multiple nucleotides (preferably between 15-30 nucleotides), (ii) flanks a polymorphic DNA target region, (iii) shares a high degree of homology among genomes of microorganisms in a taxonomic group of microorganisms, thus differentiating between organisms of certain taxonomic groups, and (iv) is able to serve as a primer binding site for PCR
20 amplification primers. In the context of this invention, a DNA region is defined as conserved when said sequence exhibits a sequence homology or nucleotide sequence similarity of at least 60%, preferably 70%, more preferably 80%, even more preferably 90% between different
25 microorganisms belonging to a single taxonomic group, wherein said sequence similarity is calculated over the entire length of the nucleic acid sequence(s).

The term “sample”, as used herein, refers to any sample suitable for analyzing or typing according to the methods of the present invention. A sample may be collected from an organism (e.g., human or other mammal,
30 plant) or environmental site (e.g. mineral, soil, rock, water). The biological

sample can be in any form, including without limitation a solid material such as a tissue, cells, a cell pellet, a cell extract, or a biopsy, or a biological fluid such as urine, blood, stool, saliva, amniotic fluid, exudate from a region of infection or inflammation, or a mouth wash containing buccal cells, urine, cerebrospinal fluid and synovial fluid and organs.

The term “amplification”, as used herein, includes methods for copying a target nucleic acid, thereby increasing the number of copies of a selected nucleic acid sequence. Amplification may be exponential or linear. A target nucleic acid may be either DNA or RNA. The sequences amplified in this manner form an “amplification product”, “amplimer” or “amplicon”, which terms are used interchangeably herein. While the exemplary methods described hereinafter relate to amplification using the polymerase chain reaction (PCR), numerous other methods are known in the art for amplification of nucleic acids (e.g., isothermal methods, rolling circle methods, etc.). The skilled artisan will understand that these other methods may be used either in place of, or together with, PCR methods. See, Saiki, “Amplification of Genomic DNA” in PCR Protocols, Innis et al., Eds., Academic Press, San Diego, Calif. 1990, pp. 13-20; Wharam et al., Nucleic Acids Res., 29(11):E54-E54, 2001; Hafner et al., Biotechniques, 30(4):852-56, 858, 860, 2001; Zhong et al., Biotechniques, 30(4):852-6, 858, 860, 2001.

The terms “amplification product”, and “amplicon”, as used interchangeably herein, refer to a nucleic acid fragment that is the product of a nucleic acid amplification or replication event, such as for instance formed in the polymerase chain reaction (PCR).

The term “template”, as used herein, refers to the nucleic acid from which the target sequence is amplified in a nucleic acid amplification reaction. The term “amplifiable template”, as used herein, refers to a template that, when amplified, results in a single amplicon. Amplifiable templates comprise primer binding sites for hybridization of amplification primers.

The term "primer" as used herein refers to a single stranded nucleotide sequence which is capable of acting as a as used herein refers to a single-stranded oligonucleotide capable of acting as a point of initiation of template-directed DNA synthesis under appropriate conditions (e.g., buffer, salt, temperature, and pH) in the presence of nucleotides and an agent for nucleic acid polymerization (e.g., a DNA-dependent or RNA-dependent polymerase). Generally, the sequence of the primer is substantially complementary to a nucleic acid strand to be copied, or at least comprises a region of complementarity sufficient for annealing to occur and extension in the 5' to 3' direction therefrom. The primer may be a DNA primer, RNA primer, or a chimeric DNA/RNA primer. Primers are preferably synthetic oligonucleotide sequences of about 12-100 nucleotides in length; preferably, about 30-60 nucleotides in length. The term "primer" may refer to more than one primer, particularly in the case where there is some ambiguity in the information regarding one or both ends of the target region to be amplified. If a "conserved" region shows significant levels of polymorphism in a population, mixtures of primers can be prepared that will amplify such sequences, or the primers can be designed to amplify even mismatched sequences. A primer can be labeled, if desired, by incorporating a label detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive labels, fluorescent labels, electron-dense reagents, enzymes (as commonly used in ELISAs), biotin, or haptens and proteins for which antisera or monoclonal antibodies are available. Preferred labels for use in this invention comprise fluorescent labels, preferably selected from FAM, TET, HEX, Cy5, Cy5.5, Cy3, Cy3.5, Cy7, Alexa dyes Tamra, ROX, JOE, FITC, and TRITC.

The term "primer set", as used herein, refers to the primer pair consisting of at least one forward primer and at least one reverse primer used in a PCR amplification reaction.

The term “primer binding site”, as used herein, refers to a specific region of the DNA fragment or segment that, as a result of its DNA sequence, is receptive of binding a PCR amplification primer having a complementary DNA sequence through DNA hybridization. A primer
5 binding site preferably ranges in size of between 15-30 nucleotides. The primer binding site comprised in the upstream adapter DNA sequence preferably differs in DNA sequence from the primer binding site comprised in the downstream adapter DNA sequence. Primer binding sites in the adapter DNA sequence preferably differ in DNA sequence from the primer
10 binding site comprised in the conserved regions of the 16S and 23S rRNA genes used for amplifying the ITS regions from the genomic DNA of the microbiome investigated

The term “co-amplified”, as used herein, refers to the simultaneous amplification of different nucleic acid fragments in a single
15 amplification reaction.

The term “target nucleic acid” as used herein refers to the nucleic acid that intended to be amplified in a nucleic acid amplification reaction, and in particular to the part of the template nucleic acid positioned between and including the primer binding sites.

20 The term “amplification efficiency”, as used herein, refers to the amount of amplification product produced in an amplification reaction from a given initial number of target sequences in a given number of amplification cycles. Thus, the amplification efficiencies of two reaction which differ only in the length of the target sequences are compared by
25 quantitatively measuring the amount of product formed in each reaction. The amplification efficiency is a measure of the efficacy of amplicon formation in a PCR reaction and can be calculated by determining the output amplicon copy number (or product) over the input template copy number. Determination of PCR amplification efficiency is well known to one

of skill in the art and is for instance explained in detail in such publications as Lalam, 2006, J Theor Biol. 242(4):947-53.

The term "standard curve", as used herein, refers to an equation or function that describes the measured relationship between the length of a DNA template and the amount of amplification product produced from this template in a PCR amplification reaction in a given number of amplification cycles.

The term "PCR calibrator", as used herein, refers to an amplifiable DNA fragment or segment serving, inter alia, as a reference template for determining the length-dependent amplification efficiency in PCR amplification reactions as described herein, in particular for determining the efficiency of the amplification reaction in amplifying sample templates of different length. At least one PCR calibrator is included in each PCR reaction. PCR calibrators consisting of *rrn* operons or parts thereof of existing microorganisms, wherein the intergenic region between the 16S and 23S rRNA genes are flanked by their native 16S and 23S rRNA gene sequences or at least by conserved regions thereof, are not part of this invention.

The term "flanked", as used herein, refers to having a defined DNA sequence (i.e. an adapter DNA sequence) contiguous with both ends of a given DNA sequence or segment (i.e. an ITS region DNA sequence).

The terms "upstream" and "downstream", as used herein, refer to a position of a genetic element on a polynucleotide sequence in relation to another genetic element. A first genetic element is upstream to a second genetic element when located in the 5' direction of the second element. A first genetic element is downstream to a second genetic element when located in the 3' direction of the second element.

The term "plasmid", as used herein, refers to a circular, double-stranded unit of DNA that replicates within a cell independently of the chromosomal DNA.

The term “replicon”, as used herein, refers to a DNA molecule or RNA molecule, or a region of DNA or RNA, that replicates from a single origin of replication. Preferably, the replicon is a plasmid.

The term “microorganism”, as used herein, refers to any
5 unicellular microorganism including bacteria, archaea, protists, fungi, virus, and algae, preferably bacteria. The term “microbial” indicates pertaining to, or characteristic of a microorganism.

The term “intestinal flora”, as used herein, refers to the population of microorganisms inhabiting the gastrointestinal tract.

10 The term "genomic DNA" as used herein refers to any DNA comprising a sequence that is normally present in the genome of a prokaryotic or eukaryotic cell or a virus. The term refers in particular to the full complement of DNA contained in the genome of a cell or organism comprising the full collective gene set of a cell. In obtaining a sample of
15 genomic DNA from a microbiome for quantitative population analysis, preferably the genomic DNA of essentially all cells in the population is isolated and such a genomic DNA sample is also referred to as total DNA. Total genomic DNA extraction procedures from divers microbiomes are known in the art and commercial genomic DNA isolation kits for this
20 purpose can be obtained from various manufacturers.

The term “population”, as used herein, refers to a plurality of individual organisms, in the context of this invention, the term refers in particular to a collection of organisms of divers taxonomic affiliation, in particular bacteria.

25 The term “diversity”, as used herein, refers to the extent to which different taxonomic groups of microorganisms are present in a population of microorganism. In order to quantify and compare microbial taxonomic diversity, i.e. “within-sample diversity”, diversity calculation using the Shannon index is recommended (Haegeman et al. 2013. ISME J
30 2013;7:1092-101). Diversity analysis, including “between sample diversity”

analysis may be performed using R software vegan package (Oksanen, J. et al. vegan: Community Ecology Package. R package version 2.0-7 (2013)). Sample diversity may be assessed for instance at the level of an individual phylum, such as on the level of the phylum *Firmicutes*, *Bacteroidetes*, or
5 *Proteobacteria*. Alternatively, the sample diversity may be assessed overall. High diversity is equivalent with a high taxonomic variation, meaning that many different bacterial taxons (e.g. different bacterial species) are represented in the population, while a low diversity is equivalent with a low taxonomic variation, meaning that a population is characterized by
10 relatively few bacterial taxons. In intestinal (fecal) flora, a Shannon index above 3 for the overall microbiota is considered to represent a high bacterial population diversity.

The term “taxonomic variation” refers to the diversity in groups of microorganisms when grouped into species, genera, families, orders, classes
15 and/or phyla in accordance with a biological classification scheme.

The term “typing”, as used herein, refers to classifying a test signature as either corresponding to the signature of a first condition or corresponding to the signature of a second condition in a classification scheme, such as a scheme classifying healthy and diseased subjects.

20 The terms “subject”, as used herein, includes reference to a human and non-human animal, preferably a human adult in the age of 18 years or older. The term “healthy subject” refers to a control subject not suffering from a disease of interest, although such a control subject may suffer from a different or related disease. The term “patient” refers to a
25 “subject” who is, or is suspected to be, afflicted with the disease of interest. In the context of Necrotising Enterocolitis (NEC), the subject or patient is preferably a prematurely born subject, more preferably a prematurely born subject having an age between 0 and 18 years, most preferably a prematurely born at a gestational age from 24 to 37 weeks. In the context of
30 sepsis, the subject or patient is preferably a neonate.

The term “neonate”, as used herein, refers to a newborn infant less than four weeks old, the term including prematurely born infants.

The term “prematurely born subject”, as used herein, refers to a subject that is born at less than 37 weeks gestational age.

5 The phrase “method for diagnosing or monitoring”, in the context of a method for diagnosing or monitoring Necrotising Enterocolitis (NEC) and/or sepsis in a subject, comprises predicting whether Necrotising Enterocolitis (NEC) and/or sepsis will develop in said subject.

 The term “signature”, as used herein, refers to a profile of
10 amplified nucleic acid fragments representing the diversity in 16S-23S ITS regions from microorganisms in a genomic DNA sample from a sample population, wherein different ITS regions represent different taxonomic groups of microorganisms, preferably said profile representing the prevalence of the various taxonomic groups of microorganisms in the sample
15 population.

 The terms “desirable signature” and “undesirable signature”, as used herein, refer in general to a beneficial signature and an unfavorable signature, these signatures representing an elected biological profile such as (i) non-diseased versus diseased, (ii) sterile versus non-sterile, (iii) non-
20 infected vs. infected, (iv) advantageous for plant growth due to the presence of a specific micro-organism vs. disadvantageous due to the absence of said specific micro-organism.

 The term “reference signature”, as used herein, refers to a signature of a control subject.

25 The term “phylum”, as used herein, refers to a taxonomic rank below kingdom and above class.

I. PCR Calibrator system

 Polymerase Chain Reaction (PCR) is a process in which DNA is
30 amplified. Inhibition of PCR can result in false negative results, and can

seriously hamper quantification of taxonomic groups in a microbiome. Inhibition can be assessed by using an internal control DNA fragment and primers specific for that fragment. If the internal control DNA fragment is not amplified, it can be deduced that the PCR reaction was inhibited. For
5 microbiome analysis it is not known how PCR inhibition affects the outcome of the analysis, in particular in view of the fact that ITS regions are polymorphic in length. Proper quantitation of initial copy number of all amplicons in samples comprising multiple amplicons (originating from multiple bacteria) requires information on potential length-specific
10 inhibition and is of great significance to e.g. testing for contamination in commercial probiotic mixtures, or for determining the abundance of *Streptococcus pneumonia* as having a normal levels or the level of infection in a throat microbiome.

Another problem in microbiome analysis, in particular when the
15 IS-pro method is performed using two or more separate PCR reactions (e.g. one for the phyla *Firmicutes/Bacteroidetes* and one for the phylum *Proteobacteria*), is that the information generated by these separate PCR reactions needs to be properly merged in order to obtain quantitative results that can be compared or merged. It would be desirable to have an indication
20 of how separate results should be transformed to render them comparable.

The length of DNA fragments generated in an IS-pro reaction cannot directly be inferred from their transition time in capillary electrophoresis. Therefore fragment length markers are routinely incorporated in the PCR reaction prior to electrophoretic analysis of the
25 amplification products. By comparing transition times of such marker fragments to the transition times of unknown microbiome fragments, the length of the unknown fragments can be inferred. This requires correct identification of the marker fragments, which process is automatically performed in the custom IS-pro pre-processing software, but this system
30 would benefit from a supplemental check for correct marker identification.

The presently proposed PCR calibrator system now provides a solution to all the above aspects of sample analysis using the IS-pro method, including (i) reaction inhibition, (ii) length-dependent amplification, (iii) calibration of electrophoresis length markers, and (iv) scaling of profiles of
5 separate IS-pro PCR reactions. This greatly improves reliability and reproducibility of the results of the IS-pro method, rendering the method uniquely suitable for, *inter alia*, research, clinical diagnostics and quality assurance.

The PCR calibrator system as disclosed herein comprises a set of
10 PCR amplification primers and a set of at least two PCR calibrators.

Preferably the set of amplification primers consists of a single forward primer and a single reverse primer for amplifying all PCR calibrators in the system. At least one of the primers in the set of calibrator primers, preferably the forward primer, comprises a label, preferably a
15 fluorescent label. Since one or more of the IS-pro primers (i.e. the taxonomic group-specific primers used to amplify the sample ITS regions and giving rise to labelled IS-pro amplicons) are in preferred embodiments also fluorescently labelled, it is essential that the fluorescent label in the set of calibrator primers differs from the label(s) of the IS-pro primers.

The PCR calibrator primers are preferably designed such that there is no interference with the actual IS-pro analysis. Hence, random amplification of non-target DNA by non-specific binding of the PCR calibrator primers in a reaction mixture comprising genomic DNA of a complex microbial population should essentially be avoided. An effective
25 PCR reaction that produces essentially no non-specific amplicons requires stable and specific binding between the 3' ends of the primer and its DNA template; in contrast, the 5' end is not critical for PCR. The uniqueness of the 3'-end subsequence of a primer mainly determines the specificity of the primer, although the binding stability is related to the entire primer
30 sequence. The T_m value and Gibbs free energy (ΔG) of the binding between

calibrator primer and its target in the adapter sequence are preferably optimized for binding stability in order to minimize non-specific amplicon production in the IS-pro reaction. Moreover, the T_m value of the calibrator primer is preferably compatible with the T_m value of the IS-pro primers.

5 The present inventors have found that using reverse sequences of the conserved regions serving as primer binding sites for the IS-pro primers provide for very suitable primer binding sites for the calibrator primers. These binding sites are facilitated by means of adapter sequences flanking the PCR calibrator fragments. In a highly preferred embodiment, the set of
10 calibrator primers in the calibrator system of this invention comprises the following primers:

MPICf 5' - GACCTAGTGGAGGAAAGATAC - 3'

MPICr 5' - TCCGTAGGTGGCACGCGGGA - 3'

wherein MPICf represents the PCR calibrator system forward primer, and MPICr represents the PCR calibrator system reverse primer.

Primer binding sites for the calibrator primers are provided in the
15 form of flanking adapter DNA sequences located upstream and downstream of the PCR calibrator fragments.

The PCR calibrator system as disclosed herein further comprises a set of at least two PCR calibrators, wherein each PCR calibrator represents a semi-artificial fragment of DNA that serves as an amplification
20 template for the calibrator primers. The PCR calibrators comprise a spacer region DNA sequence (preferably representing a microbial 16S–23S rRNA ITS region) of a given length wherein the spacer region DNA sequence comprised in each of said PCR calibrators in said set of at least two PCR calibrators is of a different length. The term “ spacer region DNA sequence
25 of a given length” in the context of the PCR calibrators refers to a DNA sequence essentially representing an amplifiable DNA sequence flanked by adaptor sequences that is co-amplified together with the DNA sequence of a 16S–23S rRNA ITS region of a microorganism. The purpose of the spacer

region in the calibrator is to provide an amplifiable sequence that is amplified with the same characteristics as the sample amplification template in an IS-pro reaction (i.e., the ITS region). Hence, it is preferred that the sequence of PCR calibrator of a given length resembles the sample

5 amplification template in an IS-pro reaction as closely as possible. It will therefore be understood that in order to prepare such a calibrator of different length, one may start from any existing DNA sequence of a microbial 16S–23S rRNA ITS region of a given length and reduce or increase the length thereof, for instance by shortening that region using

10 restriction enzymes or other molecular biological techniques, or by ligating that region to other ITS regions or parts thereof. Alternatively, an artificial or semi-artificial DNA sequence can be used as the spacer region sequence in the PCR calibrators of this invention, the exact sequence of which is not essential. In preferred embodiments, the PCR calibrator comprises as the

15 spacer region at least a part of a 16S–23S rRNA ITS region of a microorganism. The term “at least a part of a 16S–23S rRNA ITS region of a microorganism” as used herein refers to a DNA sequence of at least 20, preferably at least 30, 40, 50, 60, 70, 100, 150, or 250 consecutive nucleotides having at least 95%, preferably at least 99%, more preferably

20 100% sequence similarity with (part of) the DNA sequence of a 16S–23S rRNA ITS region of a microorganism selected from the phyla *Firmicutes*, *Fusobacterium*, *Deferribacteres*, *Spirochaetes*, *Cyanobacteria*, *Acidobacteria*, *Nitrospina*, *Nitrospirae*, *Caldithrix*, *Haloanaerobiales*, *Verrrucomicrobia*, *Chlamydiae*, *Planctomycetes*, *Gemmimonas*, *Fibrobacteres*, *Chlorobi*,

25 *Bacteroidetes*, *Proteobacteria*, *Thermotogae*, *Corprothermobacter*, *Synergites*, *Thermodesulfobacteria*, *Desulfurobacterium*, *Aquificae*, *Deinococcus-Thermus*, *Chloroflexi*, *Tenericutes* and *Actinobacteria*, preferably from the phyla *Bacteroidetes*, *Firmicutes*, *Actinobacteria*, *Proteobacteria*, *Fusobacteria* and *Verrrucomicrobia*.

In a preferred embodiment, the PCR calibrator system consists of a set of at least 3, 4, more preferably about 5, 6, 7, 8, 9 or 10 PCR calibrators of different length. The various PCR calibrators in the PCR calibrator system preferably all have different lengths, which lengths are more or less equally distributed over the expected length spectrum of the IS-pro amplimers. Preferred lengths for the PCR calibrators in a system of this invention result in calibrator amplicons between 50 and 1200 base pairs, preferably, a 5-calibrator set can consist of calibrators of around 150, 300, 450, 650 and 850 base pairs.

The PCR calibrator system may serve one of multiple purposes as explained above. Although the PCR calibrator system is preferably added to each separate IS-pro PCR reaction that is performed in the course of a single composition analysis, the PCR calibrator system for the purpose of providing a standard curve as defined herein can be used in a parallel PCR reaction, that is performed separately from the taxon-specific IS-pro PCR reaction(s). It is preferred in such cases that the parallel reaction is performed in the presence of a sample of genomic DNA of the microbiome that is studied, because this provides for conditions that allow for the occurrence of the same PCR inhibiting effect for which the IS-pro PCR reaction is to be corrected. If such a parallel reaction is used, it is also essential that at least one PCR calibrator is present in the IS-pro PCR reaction(s) so as to provide an internal control, the amplification efficiency of which can be compared with the amplification efficiency of the PCR calibrator having a corresponding length in the parallel reaction for the standard curve. In this case, both length-dependent correction of amplification efficiency and scaling of amplitude of the amplification (the overall efficiency) can be achieved.

In preferred embodiments, the PCR calibrator system is incorporated in the IS-pro reaction and the PCR calibrators are co-amplified with the sample ITS regions. When amplified in the IS-pro reaction (or in a

parallel reaction as the case may be), the PCR calibrator system results in a specific peak profile (see Figure 1A). This peak profile can be used to perform one or more corrective calculations on the amplification results of the IS-pro reaction.

5 In a preferred embodiment, each of the at least two PCR calibrators in the set is present in exact equal amounts, although each PCR calibrator may also be present in known amounts relative to any other PCR calibrator in said set. This eases the quantitative analysis of the PCR reaction results. In order to prepare a mixture of PCR calibrators present in
10 equal amounts, the present inventors have found that a copy of each of the calibrator sequences, each including its flanking adaptor sequences, can be inserted in a plasmid, and the plasmid can be cloned and replicated in a suitable host. The plasmid DNA can then be isolated and cut by restriction enzymes having restriction sites located between each PCR calibrator,
15 thereby providing a preparation of PCR calibrators having different length and present in equal amounts.

A set of at least two PCR calibrators as described herein above, wherein said set is comprised in a single replicon or single amplifiable template, preferably a DNA plasmid, is an aspect of this invention.

20

Inhibition of the PCR reaction

An advantage of the proposed calibrator system is that it can provide information on inhibition. Inhibition of the PCR reaction can be identified when the PCR calibrator system does not yield a profile.

25 Moreover, if the profile is lower than the PCR calibrator profile that is found in a parallel reaction or in a blank sample (with no target DNA for the regular PCR reaction), it can be assumed that there is inhibition to a certain extent which can be corrected for. This correction may be done by up-scaling the PCR calibrator profile to the same level (peak heights) as found in the
30 blank sample. The same up-scaling factor can then be applied to all other

peaks present in the profile (Figure 2). Finally, a cut-off can be defined for the level of inhibition, based upon which a sample can be accepted or refused for further analysis by an (automated) quality system.

Alternatively, if a single internal control PCR calibrator as defined herein is used on concert with a parallel PCR calibrator system for generating the standard curve, an up-scaling factor may also be calculated based on the internal control PCR calibrator that can then be applied to all peaks of the IS-pro profile, while the parallel PCR calibrator system consisting of at least two PCR calibrators of different length amplified in a genomic DNA sample having the same potential inhibiting effect could provide information on the amplification efficiency of the various peaks present in the profile that represent microbial ITS regions of different length.

Amplification efficiency of small vs large fragments

Another advantage of the proposed calibrator system is that it can be used to identify differences in amplification efficiency between short and long fragments of DNA. Because all DNA fragments in the PCR calibrator system are present in exactly the same amount (or in defined amounts relative to each other, depending on the embodiment chosen), the peaks that result from fragment analysis after PCR should all be equal in height (or have defined heights relative to each other). However, as shown in Figure 1A, this may not be the case in practice. The differences in peak heights are caused by dissimilar amplification efficiencies. As can be seen in Figure 1A, this effect is generally not linear (a straight line connecting the tops of all peaks cannot be drawn). In case that the PCR calibrator system includes more than two fragments (as is the case in the embodiment of Figure 1), the overall (nonlinear) function of amplification efficiency over fragment length can be estimated. This can be done by means of a multi-linear model, in which peak tops are connected by means of straight lines (Figure 1B), or by a more sophisticated nonlinear model (Figure 1C). Finally, the estimated

amplification efficiency function can be used to apply a corrective factor to all peaks present in a profile (Figure 1D). Note: This whole procedure can be combined with a corrective up-scaling factor when samples are partially inhibited as mentioned above.

5

Error checking the pre-processing procedure

Another advantage of the proposed calibrator system is that it can be used to verify correct fragment length determination in the capillary electrophoresis software. To infer fragment lengths from their transition
10 time in capillary gel electrophoresis, a set of fragments of known length is added to the process (so-called length markers). These fragments can be readily seen in the raw capillary gel electrophoresis data (Figure 3A). Such data are subsequently analyzed by custom pre-processing software. This software tool identifies every fragment belonging to the length marker and
15 assigns a length value to it. It is important to supervise this process independently, as it may have an impact on all subsequent analyses. For this purpose, the presently proposed PCR calibrator system can be used. The PCR calibrator system, the length marker and the IS-pro fragments are in a preferred embodiment all run concurrently in the same capillary in the
20 capillary gel electrophoresis machine. Information of fragment lengths of the length marker can in such instances be used to infer lengths of the other fragments. The PCR calibrator system, internal control and IS-pro fragments, in a preferred embodiment, all have different fluorescent labels, allowing their separate analysis. To evaluate if any errors have occurred in
25 the identification of the length marker peaks during the pre-processing step, the position of the peaks of the PCR calibrator system as detected in the same capillary electrophoresis run can be used. After identification of length marker peaks, all other peaks (from PCR calibrator system and IS-pro) can be assigned length values. As the PCR calibrator peaks have known lengths,
30 they can be used to control the identification of the length marker peaks. To

do this, the lengths of the PCR calibrator peaks as assigned by the software tool are compared to their real (known) lengths. If these values correspond, it may be safely assumed that the length marker peaks have been identified correctly.

5

Scaling of profiles from separate PCR reactions

As stated above, the IS-pro method may, in preferred embodiments, consist of two or more separate PCR reactions for amplifying different taxonomic groups of microorganisms. For some applications of the IS-pro method it is necessary to combine information from these two or more separate PCRs in a quantitative manner, e.g. when abundances of all taxonomic groups in a complex sample are to be investigated. To do this, a reference signal is needed that is the same in both reactions. For this purpose, the PCR calibrator system can be used. By comparing the height of the PCR calibrator peaks in one PCR reaction with their corresponding height in the other reaction, and knowing the amount of calibrator added to the mixtures, quantitatively comparative results can be obtained by up- or downscaling all peaks in one of the samples in the same fashion as was described for partially inhibited samples herein above. After this correction, peak profiles can be compared on a quantitative level.

10

15

20

The advantages of the PCR calibrator system as disclosed herein will become more apparent from the Examples below.

II. IS-pro analysis

25

The present invention provides in another aspect a method of analyzing the composition of a microbiome based on taxonomic variation in the DNA sequence of the microbial 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the microorganisms in said microbiome.

The method is previously described as the IS-pro method in Budding et al. 2010, and an exemplary method is described in great detail in Example 2, below. IS-pro involves bacterial species differentiation by the length of the 16S–23S rDNA interspace region with taxonomic classification
5 by phylum-specific fluorescent labelling of PCR primers.

Amplifying the 16S–23S rRNA intergenic spacer region from microorganisms is well known in the art. In order to achieve this, the sequences of conserved DNA regions comprised in the 16S and 23S rRNA gene sequences flanking the intergenic region in the genomic DNA of the
10 microorganism are used as primer binding sites for amplification of the polymorphic DNA region on which the taxonomic diversity analysis of IS-pro is based.

Prokaryotic microorganisms, including bacteria and archaea, comprise in their genome one or more copies of the rrn operon comprising
15 the genes for the 5S, 16S and 23S ribosomal RNAs. In most prokaryotes the ribosomal genes in the operon are in the order 16S-23S-5S and are co-transcribed in a single polycistronic RNA that is processed to provide the RNA species present in the mature ribosome. The rRNA 16S and 23S genes have acquired paramount relevance for the study of bacterial evolution and
20 phylogeny, and the presence of variable and conserved regions in both genes is well documented (Neefs et al., 1993. *Nucleic Acids Research* 21(13): 3025-3049; Baker et al. 2003. *J. Microbiol. Meth.* 55: 41– 555; Van de Peer et al., 1996. *Nucleic Acids Research* 24(17): 3381–3391; Gurtler and Stanisich, 1996. *Microbiology* 142: 3-16). The spacer between the 16S and 23S genes
25 contains regions with secondary structures and sometimes tRNA genes. The variation found among relatively close taxa is known to be very high for the spacers of the rRNA operons. The extreme divergence in size and sequence of the spacers among different groups of prokaryotes, together with their location between highly conserved rRNA genes, makes them ideally suited
30 as taxonomic markers. The 16S–23S rRNA intergenic spacer region is

amplified using primers directed to conserved regions in the ribosomal gene sequences. More preferably, the conserved DNA regions are those located nearest to the 3'-end of the 16S rRNA gene and nearest to the 5'-end of the 23S rRNA gene.

5 Amplification of ribosomal sequences is described in detail in Devereux and Wilkinson 2004 (Molecular Microbial Ecology Manual, Second Edition 3.01: 509–522, 2004). The pivotal point for the purpose of microbiome analysis using phylum-specific probes is the primer design. Primers can be designed that selectively amplify rRNA genes of
10 phylogenetically defined groups. Selection of primers can often be guided by comparison of sequences in a database. Many selective rRNA gene primers have been described in the literature. PRIMROSE (Ashelford et al. 2002. Nucl Acids Res 30: 3481–3489) is a program that uses sequences from the Ribosomal Database Project to identify and determine the phylogenetic
15 range of oligonucleotides that may be used as rRNA probes or primers. ProbeBase (Loy et al. 2003. Nucl Acids Res. 31, 514–516) is a database of published rRNA probes with information on target site and specificity. The Ribosomal Database Project II (Maidak et al. 2001. Nucl Acids Res 29:173–174) and the ARB package (Ludwig et al. 2004. Nucleic Acids Research
20 32(4):1363-1371) provide software for *in silico* evaluation of intended specificities of ribosomal primers, including those for 16S and 23S rRNAs, against known rRNA sequences.

 The present inventors have found that the amplification of the phylum of *Proteobacteria* is very beneficially performed by carrying out a
25 multiplex PCR in order to provide sufficient taxonomic resolution within the proteobacterial phylum. By simultaneously amplifying multiple sequences in a single reaction, a process referred to as multiplex polymerase chain reaction (PCR), deep resolution can be obtained. Although alternative proteobacterial primers may be used in embodiments of aspects of this

invention, the following primer set is preferably used in several embodiments as indicated herein:

Primers *Proteobacteria*

Forward:

ProtISf 5' - CCGCCCGTCACACCATGG - 3'

Reverse:

DPISr1 5' - AATCTCGGTTGATTTCTTTTCCT - 3'

DPISr2 5' - AATCTCGGTTGATTTCTTCTCCT - 3'

DPISr3 5' - AATCTCTTTTGATTTCTTTTCCTCG - 3'

DPISr4 5' - AATCTCATTTGATGTCTTTTCCTCG - 3'

DPISr5 5' - AATCTCTTTTGATTTCTTTTCCTTCG - 3'

DPISr6 5' - AATCTCTCTTGATTTCTTTTCCTTCG - 3'

DPISr7 5' - AATCTCAATTGATTTCTTTTCCTAAGG - 3'

- 5 This primer set in any combination of the forward primer together with at least two of the reverse primers, is an aspect of this invention. Hence in one aspect, the present invention provides a set of oligonucleotide primers for amplifying sequences of the 16S-23S intergenic region of the phylum *Proteobacteria* from a sample of microbial genomic DNA, said set
- 10 comprising the forward primer indicated herein above as ProtISf in combination with at least 2 reverse primers selected from the group consisting of reverse primers indicated herein above as DPISr1, DPISr2, DPISr3, DPISr4, DPISr5, DPISr6, and DPISr7. In a preferred embodiment, at least one of said primers comprises a detectable label.
- 15 Although the present invention is described in the context of the 16S-23S rRNA intergenic spacer, the skilled person will understand that variations to other intergenic spacer regions, including intergenic spacer regions in eukaryotic rRNA operons are within the ambit of this invention. In fact, any microorganism taxonomically distinguishable on the basis of a
- 20 polymorphic DNA region flanked by conserved DNA regions, homologous

sequences of which are present in all microorganisms may be subjected to analysis to using methods of this invention.

Preferred microorganisms in aspects of this invention belong to bacterial phyla selected from the group consisting of *Firmicutes*,
 5 *Fusobacterium*, *Deferribacteres*, *Spirochaetes*, *Cyanobacteria*, *Acidobacteria*,
Nitrospina, *Nitrospirae*, *Caldithrix*, *Haloanaerobiales*, *Verrucomicrobia*,
Chlamydiae, *Planctomycetes*, *Gemmimonas*, *Fibrobacteres*, *Chlorobi*,
Bacteroidetes, *Proteobacteria*, *Thermotogae*, *Corprothermobacter*, *Synergites*,
Thermodesulfobacteria, *Desulfurobacterium*, *Aquificae*, *Deinococcus*-
 10 *Thermus*, *Chloroflexi*, *Tenericutes* and *Actinobacteria*. More preferred phyla
 targeted in methods of this invention comprise *Bacteroidetes*, *Firmicutes*,
Actinobacteria, *Proteobacteria*, *Fusobacteria* and *Verrucomicrobia*. Highly
 preferred are *Bacteroidetes*, *Firmicutes*, *Actinobacteria* and *Proteobacteria*.
 Any microbiome comprising prokaryotes belonging to these phyla
 15 constitutes a preferred microbiome in aspects of this invention.

These phyla are known to the person skilled in the art and have been described, *inter alia*, in Schloss 2004 (Microb. Mol. Biol. Rev. 6 (4): 686-691) , or in the Bergey Manual, Second Edition 2004, Release 5.0.

The microbiomes may be present in any environment.
 20 Environments according to the present invention include for example
 environments present in humans, plants, animals, water, food (like dairy
 products), yeast cultures (e.g. used in industry) or soil. More specifically,
 environments such as the (gastro)intestinal tract, vaginal tract, skin, lung,
 sputum, colon, mouth, teeth pockets, ascetic fluid, feces, purulence, abscess,
 25 wound fluid, wound, blood, or a cardiovascular system are envisaged.

The sampling of the above environments, for the purpose of obtaining a representative sample for analysis, may be performed using any suitable method available to one of skill in the art. For example, a sputum specimen may be obtained when the population of microorganisms in the
 30 lungs is to be analyzed. A sample of intestinal flora can for example be

obtained in the form of a fecal sample, a biopsy of the gastrointestinal mucosa, or a rectal swab. The present inventors have found that the use of a rectal swab provides more consistent results than other methods of sampling the intestinal flora using the IS-pro technique. Hence, an aspect of this invention is a method of performing a sampling of the intestinal flora for IS-pro analysis using a rectal swab, preferably without bead beating.

DNA samples for taxonomic composition analysis with diagnostic purpose from any environment may be obtained by using generally known techniques for DNA isolation. The total genomic DNA may be purified from study environments or from sampling devices probed in said environments by using, for instance, a combination of physical and chemical methods. Very suitably commercially available systems for DNA isolation are used, such as the NucliSENS® easyMAG® nucleic acid extraction system (bioMérieux, Marcy l'Etoile, France) or the MagNA Pure 96 System (Roche Diagnostics GmbH, Mannheim, Germany).

The genomic DNA sample is analyzed by a PCR-based profiling technique referred to herein as IS-pro as described in Budding et al., 2010 (cited above). In IS-pro, distinct 16S-23S ITS DNA sequences of different length amplified from the sample of genomic DNA are separated by capillary gel electrophoresis, providing a profile of distinct peaks each representing a separate 16S-23S ITS DNA sequence having a characteristic length (in number of nucleotides or base pairs). Each peak is considered to represent a single ITS region. Since genomes may comprise multiple rrn operons with different ITS region sequence, not every peak represents a distinct species of microbe. Hence, peaks are designated as operational taxonomic units (OTUs). In order to improve resolution and separate more peaks from the population, phylum-specific ITS regions can be recognized by using phylum-specific fluorescent labeling. Analysis of the profiles results in the identification of dominant OTUs within phyla and enables identification of variation within these phyla. The person skilled in the art will

understand that conserved DNA regions of microorganisms in a taxonomic group other than phylum can be labeled using primers containing said label. Suitable labels are described herein above.

It is preferred, but not essential to aspects of this invention that the PCR calibrator system as described herein is used as part of the IS-pro method. Although the PCR calibrator system of this invention is beneficial to the reliability and reproducibility of the method, useful analytical and diagnostic results are obtained in many analytic and diagnostic embodiments described herein without the use of the calibrator system.

Methods of the present invention may be used to type, diagnose, investigate, analyse and monitor such diverse microbiomes as those associated with the gut or gastrointestinal tract, the skin, the urogenital tract, the oral cavity, and the pulmonary system, and diagnose, monitor or predict disease. The methods may be used for diagnostic purpose such as for diagnosing, monitoring or predicting (incl. early detection of) a microbial infection, and may even be used for environmental diagnostics, such as for determining the microbial status of a sample source, wherein said sample source is of environmental, plant, animal or food origin, or a sample of a pharmaceutical or chemical product intended to be devoid of microbes or microbial DNA and wherein specific profiles of ITS regions (or the absence thereof) are indicative of sterility of said sample, quality of the environment, microbial safety of a food, pharmaceutical product, the quality of a chemical product or the health of plant or animal.

III. Machine learning algorithms for predicting clinical status of samples

Diagnosis, monitoring or typing can be performed by assessing the absence or presence of specific peaks, the abundance thereof, or the peak volume, or by applying machine learning algorithms on at least one, preferably at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 signatures or profiles, preferably having been typed previously as desirable (e.g. healthy) and/or undesirable (e.g. diseased) signatures or profiles. Suitable machine learning algorithms

include Support Vector Machines, Random Forest or partial least squares discriminant analysis (the result of the latter is for example displayed herein as ROC curves). The skilled person is aware of these and other machine learning programs and knows how to apply them on generated data sets in order provide a tool for diagnosis, monitoring or typing in methods as described herein. One machine learning algorithm, the partial least squares discriminant analysis (PLS-DA), is described herein below.

Partial least squares discriminant analysis (PLS-DA)

A partial least squares discriminant analysis (PLS-DA) regression model can be used for predicting whether a target or test sample belongs to a desirable signature or to a non-desirable signature; for example whether it belongs to a diseased subject or to a healthy subject. PLS-DA is a supervised classification method, which aims to find linear transformations of a matrix of predictors and categorical responses so as to maximize their covariance. [Rajilić-Stojanović M, Biagi E, Heilig HG, et al. 2011. Gastroenterology 141:1792-801]. The PLS-DA model can be constructed on the basis of four different datasets: one for each of the three separate phylum groups and one for the overall microbial composition, by pooling all phyla. Under the assumption that the more discriminant variables are the ones with a higher variance, a preliminary variable selection can be performed by filtering out low variance predictors.

PLS-DA model validation can be carried out by a 10-fold cross-validation procedure. In practice, the dataset is split into 90% of samples for model construction (i.e. the training set) with the aim to predict the other 10% (i.e. the test set). This procedure is repeated for 10 iterations, where each sample serves as a test sample exactly once. Accuracy rates, specificity and sensitivity are computed for the samples that are used as a test set in every iteration, and the model predictive power is further assessed using a

receiver operating characteristic (ROC) curve, a function of the true positive rate (TPR or sensitivity) and false positive rate (FPR or 1-specificity).

PLS-DA provides a quantitative estimate of the discriminatory power of each descriptor by means of VIP (variable importance for the projection) parameters. VIP values rank the descriptors by their ability to discriminate different groups and are therefore considered an appropriate quantitative statistical parameter. The VIP criterion can be used to rank the different OTUs based on their contribution to the response variable (for example, clinical status such as diverticulitis: present or absent) and PLS components. Only the OTUs with the highest contribution (VIP score > 1.2) are considered. The OTUs resulting from this selection are translated to most likely bacterial species by comparison to a database consisting of >1500 bacterial species and their associated IS lengths. PLS-DA analysis can be performed using the DiscrMiner package in R (version 2.15.2). All data visualizations can be performed with the Spotfire software package (TIBCO, Palo Alto, CA, USA).

IV. The present invention also envisions that a sample of interest is collected and provided by a user, such as a patient or medical practitioner, to a testing facility for determining a signature of the microbiome as described herein. The methods of the present invention may then be performed by the testing facility on the sample provided by the user, and the resulting test signature may subsequently be returned by the testing facility to the user that has provided the test sample.

The present invention furthermore envisions a system wherein a user in possession of a test signature, is able to provide the test signature to an analysis facility for providing a diagnostic test result based on the test signature. The analysis facility is preferably a computer system comprising one or more of a database of reference signatures stored on a computer-readable storage medium, a computer program comprising instructions configured to cause a computer system to perform the operations of

comparing test signatures with reference signatures, and machine learning algorithms for predicting the diagnostic status of signatures based on comparison with a learned set of reference signatures, such as for predicting the diagnostic test result of a clinical sample, for instance as described in
5 great detail herein above, and in the Examples herein below. The test signature may be provided to the analysis facility in the form of an electronic input message, optionally following a user authentication process, such as a verification process based on a pin code or biometric code.

The computer system of the analysis facility generates an
10 electronic output message containing the predicted diagnostic test result and this message may be stored on a computer-readable storage medium, and/or the message may be communicated to a user-specified device, such as a smartphone or tablet. The communication between the computer system and the user-specified device may be via a wireless network system, the
15 internet, email, sms, WhatsApp, or dedicated apps, and access to the contents of the output message is optionally again subject to secure access using a user authentication process, such as a verification process based on a pin code or biometric code.

20 Embodiments of this invention are *inter alia*:

Embodiment 1. A PCR calibrator system, comprising a set of PCR amplification primers at least one of which primers comprises a label, and a set of at least two PCR calibrators, each PCR calibrator consisting of a DNA fragment comprising a spacer region having a DNA sequence of a given
25 length flanked by upstream and downstream adapter DNA sequences that comprise primer binding sites for binding of said PCR amplification primers wherein said set of PCR amplification primers is for PCR amplifying the spacer region DNA sequence of all PCR calibrators in said set of at least two PCR calibrators, wherein the spacer region DNA sequence comprised in
30 each of said PCR calibrators in said set of at least two PCR calibrators is of

a different length, and wherein each PCR calibrator in said set of at least two PCR calibrators is present in equal amount or in a known amount relative to other PCR calibrators in said set.

5 Embodiment 2. A PCR calibrator system according to embodiment 1, wherein said spacer region DNA sequence is the sequence of at least a part of a microbial 16S–23S rRNA internal transcribed spacer (ITS) region, and wherein said adapter DNA sequences in said set of at least two PCR calibrators replace the DNA sequences of the conserved DNA regions
10 comprised in the 16S and 23S rRNA gene sequences upstream and downstream of said ITS region in the microbial genome from which said ITS region originates and wherein said adapter DNA sequences have less than 30% sequence identity with the DNA sequences of said conserved DNA regions.

15

Embodiment 3. A set of at least two PCR calibrators as defined in embodiment 1 or 2, wherein said set is comprised in a single replicon or single amplifiable template, preferably a DNA plasmid.

20 Embodiment 4. A method of analyzing the composition of a microbiome based on taxonomic variation in the DNA sequence of the microbial 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the microorganisms in said microbiome, wherein the sequences of conserved DNA regions comprised in the 16S and 23S rRNA sequences flanking said
25 ITS region in the genomic DNA of said microorganisms comprise primer binding sites for amplification of said ITS regions, said method comprising the steps of:

- a) providing a sample of genomic DNA from a microbiome;
- b) providing a PCR calibrator system as defined in embodiment 1 or 2;

- c) adding said set of at least two PCR calibrators from said PCR calibrator system to said sample of genomic DNA;
- d) performing a PCR amplification reaction on said sample of genomic DNA comprising said set of at least two PCR calibrators using said set of PCR
- 5 amplification primers from said PCR calibrator system as a first set of amplification primers to amplify and provide amplification products of said ITS region(s) comprised in said set of at least two PCR calibrators, and using at least a second set of PCR amplification primers directed to said flanking conserved DNA regions to thereby co-amplify and provide
- 10 amplification products of said ITS regions comprised in said sample of genomic DNA, and;
- e) providing a standard curve by determining the PCR amplification efficiency of each of said at least two PCR calibrators from said PCR calibrator system in said PCR amplification reaction of step d) and
- 15 expressing said PCR amplification efficiency as a function of the length of the DNA sequence of the ITS region;
- f) determining the length-specific amplification efficiency for ITS regions of different length comprised in said genomic DNA sample and amplified in step d) using the standard curve as provided in step e);
- 20 g) determining the abundance of microbial 16S–23S rRNA internal transcribed spacer (ITS) regions of different length in said microbiome using the length-specific amplification efficiencies determined in step f), and
- h) analyzing the composition of a population of microorganisms based on the abundances of ITS regions of different length determined in step g).

25

Embodiment 5. The method of embodiment 4, wherein said standard curve is based on at least five PCR calibrators of different length ranging in length from 50 to 1200 bps.

- Embodiment 6. The method of embodiment 4 or 5, wherein said step d) of performing a PCR amplification reaction on said sample of genomic DNA using at least a second set of PCR amplification primers directed to said flanking conserved DNA regions comprises the use of a set of PCR
- 5 amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phyla *Firmicutes* and *Bacteroidetes* and a set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phylum *Proteobacteria*.
- 10 Embodiment 7. The method of embodiment 6, wherein said set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phyla *Firmicutes* and *Bacteroidetes* comprises the use of:
- 15 a) the forward primer 5'-CTGGATCACCTCCTTTCTAWG-3' comprising a first fluorescent label,
- b) the forward primer 5'-CTGGAACACCTCCTTTCTGGA-3' comprising a second fluorescent label;
- c) and three unlabeled reverse primers 5'-
- 20 AGGCATCCACCGTGCGCCCT-3'; 5'-AGGCATTCACCRGTGCGCCCT-3'; and 5'-AGGCATCCRCCATGCGCCCT-3'.

- Embodiment 8. The method of embodiment 6 or 7, wherein said set of PCR amplification primers for amplifying the 16S–23S rRNA internal
- 25 transcribed spacer (ITS) regions in the genomic DNA of the phylum *Proteobacteria* comprises the use of:
- a) the forward primer 5'-CCGCCCCGTCACACCATGG-3'
- b) at least one of the reverse primers selected from the group consisting of
- 5'-AATCTCGGTTGATTTCTTTTCCT-3',
- 30 5'-AATCTCGGTTGATTTCTTCTCCT-3',

- 5'-AATCTCTTTTGATTTCTTTTCCTCG-3',
 5'-AATCTCATTTGATGTCTTTTCCTCG-3',
 5'-AATCTCTTTTGATTTCTTTTCCTTCG-3',
 5'-AATCTCTCTTGATTTCTTTTCCTTCG-3',
 5 5'-AATCTCAATTGATTTCTTTTCCTAAGG-3',

wherein at least one of said primers comprises a fluorescent label.

Embodiment 9. The method of any one of embodiments 4-8, wherein said microbiome is the microbiome of the respiratory tract, oral cavity, skin,
 10 gastrointestinal tract, or urogenital tract or that of urine, saliva, sputum, pus, wound fluid, or feces, preferably of the human body, or the microbiome associated with soil, waterbodies or plants.

Embodiment 10. A method for diagnosing or monitoring a digestive tract or
 15 gastrointestinal disorder in a patient, wherein the disorder is selected from the group consisting of digestive tract or gastrointestinal disorders associated with inflammatory bowel disease (IBD), diverticulitis, irritable bowel syndrome (IBS), coeliac, lactose intolerance, Necrotising Enterocolitis (NEC), Clostridium Difficile Associated Diarrhea and colorectal cancer, or
 20 comorbid disorder selected from the group consisting of attention-deficit/hyperactivity disorder (ADHD), obsessive-compulsive disorder (OCD), anxiety, stress, eating disorder, major depressive disorder (MDD), bipolar disorder, depression, and schizophrenia, said method comprising performing the method of any one of embodiments 4-9 on a sample of the
 25 intestinal flora, and optionally treating the patient by pharmacological intervention.

Embodiment 11. Method according to embodiment 10, wherein said sample of intestinal flora is a sample obtained from small or large intestine,
 30 preferably colon or rectum, preferably in the form of a biopsy specimen of

gastrointestinal mucosa or in the form of a feces sample, more preferably wherein said sample is obtained using a rectal swab.

Embodiment 12. Method according to embodiment 10 or 11, wherein IBD
5 is selected from Crohn's disease, colitis ulcerosa, collagenous colitis, lymphocytic colitis, ischemic colitis, microscopic colitis, Behçet's syndrome, diversion colitis, diverticular colitis, eosinophilic colitis, and radiation colitis.

10 Embodiment 13. Method according to any one of embodiments 10-12, wherein the digestive tract or gastrointestinal disorder is treated by pharmacological intervention with therapeutic agent selected from anti-inflammatory agents, and antibiotic agents, TNF-alfa inhibitors, including infliximab and adalimumab, and 5-ASA formulations, including
15 sulfasalazine and mesalazine.

Embodiment 14. Method according to any one of embodiments 10-12, wherein the comorbid disorder is attention-deficit/hyperactivity disorder (ADHD) and the pharmacological intervention comprises treating the
20 patient with a methylphenidate stimulant, preferably selected from the group consisting of Adderall, Concerta, Vyvanse, and Ritalin.

Embodiment 15. Method according to any one of embodiments 10-12, wherein the comorbid disorder is selected from obsessive-compulsive
25 disorder (OCD), anxiety, stress, major depressive disorder (MDD), bipolar disorder symptoms, depression, and schizophrenia, and the pharmacological intervention comprises treating the patient with an antidepressant, preferably selective serotonin reuptake inhibitors (SSRIs), more preferably fluoxetine, citalopram, escitalopram, paroxetine, or sertraline; or an
30 antipsychotic selected from the group consisting of aripiprazole, asenapine,

clozapine, iloperidone, olanzapine, paliperidone, quetiapine, risperidone, and ziprasidone.

Embodiment 16. Method according to any one of embodiments 10-12,
5 wherein the eating disorder is anorexia nervosa or bulimia nervosa, and the pharmacological intervention comprises the administration of an antidepressant, preferably selective serotonin reuptake inhibitors (SSRIs), more preferably fluoxetine, citalopram, escitalopram, paroxetine, or sertraline; an antihistamine; or an antipsychotic, preferably selected from
10 the group consisting of aripiprazole, asenapine, clozapine, iloperidone, olanzapine, paliperidone, quetiapine, risperidone, and ziprasidone.

Embodiment 17. A method for diagnosing or monitoring a systemic disease in a patient, said method comprising performing the method of any one of
15 embodiments 4-9 on a body sample of said patient, and optionally treating the patient by pharmacological intervention.

Embodiment 18. Method according to embodiment 17, wherein said systemic disease is selected from the group consisting of arthritis,
20 sarcoidosis, mixed connective tissue disease, spondylitis ankylopoetica, osteoporosis, juvenile idiopathic arthritis, osteoarthritis, rheumatoid arthritis, sepsis, Behçet's disease, Sjögren's syndrome, fibromyalgia, sclerodermia, Raynaud's phenomenon, and systemic lupus erythematosus, said method comprising performing the method of any one of embodiments
25 4-9 on a sample of the intestinal flora.

Embodiment 19. A method for diagnosing or monitoring a skin disorder or chronic wound in a patient, said method comprising performing the method of any one of embodiments 4-9 on a sample of the skin or wound flora, and
30 optionally treating the patient by pharmacological intervention.

Embodiment 20. Method according to embodiment 19, wherein said skin disorder is psoriasis, eczema, acne, or rosacea, or wherein said chronic wound is a diabetic ulcer, or ulcer associated with vascular disease and
5 wherein the pharmacological intervention comprises the administration of an antibiotic or anti-inflammatory agent, preferably a topical corticosteroid selected from cortisone, hydrocortisone, prednisone and prednisolone, methylprednisolone, dexamethasone, triamcinolone, mometasone, fluticasone, betamethasone, halometasone and desonide, preferably wherein
10 the antibiotic agent is provided in the form of an antibiotic ointment, preferably selected from fusidin cream, flammazin cream, bactroban ointment, baneocin ointment, erythromycin unguent, chlorhexidin cream, and mupirocin cream.

15 Embodiment 21. A method of determining fertility or the probability of success of an *in vitro* fertilization or embryo transfer procedure in a female, said method comprising performing the method of any one of embodiments 4-9 on a sample of the vaginal flora, and optionally treating the patient by pharmacological intervention.

20 Embodiment 22. A method for diagnosing or monitoring bacterial vaginosis, said method comprising performing the method of any one of embodiments 4-9 on a sample of the vaginal flora, and optionally treating the patient by pharmacological intervention, preferably wherein the
25 pharmacological intervention comprises the systemic or topical administration of an antibiotic, preferably metronidazole, clindamycin, amoxicillin-clavulanate or fluconazole.

Embodiment 23. A method for diagnosing or monitoring an oral, nasal or
30 oropharyngeal disorder selected from periodontitis, periimplantitis, and

oro-nasopharyngeal carcinoma said method comprising performing the method of any one of embodiments 4-9 on a sample of the oral flora, and optionally treating the patient by pharmacological intervention, preferably wherein the pharmacological intervention comprises the systemic, topical or
5 local administration of an antibiotic or antineoplastic agent.

Embodiment 24. A method for diagnosing or monitoring a disorder that affects the upper or lower respiratory tract, selected from respiratory infection, chronic obstructive pulmonary disease (COPD), asthma, cystic
10 fibrosis and lung cancer, said method comprising performing the method of any one of embodiments 4-9 on a pulmonary sample selected from bronchoalveolar lavage, a sputum sample and a lung biopsy, and optionally treating the patient by pharmacological intervention, preferably wherein the pharmacological intervention comprises the systemic, topical or local
15 administration of an antibiotic agent, anti-inflammatory agent or antineoplastic agent.

Embodiment 25. A method for diagnosing or monitoring a microbial infection, said method comprising performing the method of any one of
20 embodiments 4-9 on a liquor sample, a pleural sample, a blood sample, a urine sample, an abscess sample, or a tissue sample of an organ, and optionally treating the patient by pharmacological intervention, preferably wherein the pharmacological intervention comprises the systemic, topical or local administration of an antibiotic agent.

25

Embodiment 26. A method for determining the microbial status of a sample, wherein said sample is of environmental, plant, animal or food origin or a sample of a pharmaceutical or chemical product intended to be devoid of microbes or microbial DNA, said method comprising performing
30 the method of any one of embodiments 4-9 on a sample from said source,

and optionally treating the source from which the sample was obtained, wherein said status is indicative of sterility of said sample, quality of the environment, microbial safety of food, or the health of plant or animal, preferably wherein the treatment comprises the reduction of microbial growth in said sample source or elimination of microbes or microbial DNA from said sample source.

Embodiment 27. A method of typing the intestinal flora of a subject for having a diverticulitis signature, comprising analyzing the composition of the population of microorganisms in said intestinal flora based on taxonomic variation in the DNA sequence of the microbial 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of said microorganisms, wherein the sequences of conserved DNA regions comprised in the 16S and 23S rRNA sequences flanking said ITS region in the microbial genome comprise primer binding sites for amplification of said ITS regions, said analysis comprising the steps of:

- a) providing a sample of genomic DNA from the population of microorganisms constituting the intestinal flora of a subject;
- b) performing a PCR amplification reaction on said sample of genomic DNA using at least one set of PCR amplification primers directed to said flanking conserved DNA regions to thereby amplify and provide amplification products of said ITS regions comprised in said genomic DNA sample;
- c) analyzing said amplification products to thereby provide a test signature of the composition of the population of microorganisms in said intestinal flora;
- d) comparing said test signature with a reference signature of a healthy subject and/or a subject suffering from diverticulitis and classifying the test signature as a signature of a healthy subject or as a signature of a subject suffering from diverticulitis.

Embodiment 28. The method of any one of embodiment 27, wherein the signature of a subject suffering from diverticulitis is characterized by increased peak area or peak height in the capillary gel electrophoresis peak profile for DNA fragments having a length in the range from 400-450
5 nucleotides, wherein said peak profile is produced by capillary electrophoretic analysis of said amplification products.

Embodiment 29. A method of typing the intestinal microbiome of a subject for having the signature of a microbiome associated with a disease selected
10 from a digestive tract or gastrointestinal disorder, or systemic disease as defined in any one of embodiments 10-18, said method comprising analyzing the composition of the population of microorganisms in the intestinal microbiome based on taxonomic variation in the DNA sequence of the microbial 16S-23S rRNA internal transcribed spacer (ITS) regions in the
15 genomic DNA of said microorganisms, wherein the sequences of conserved DNA regions comprised in the 16S and 23S rRNA sequences flanking said ITS region in the microbial genome comprise primer binding sites for amplification of said ITS regions, said analysis comprising the steps of:

- a) providing a sample of genomic DNA from the population of
20 microorganisms constituting the intestinal microbiome of a subject;
- b) performing a PCR amplification reaction on said sample of genomic DNA using at least one set of PCR amplification primers directed to said flanking conserved DNA regions to thereby amplify and provide amplification products of said ITS regions comprised in said genomic DNA sample;
- 25 c) analyzing said amplification products to thereby provide a test signature of the composition of the population of microorganisms in said intestinal microbiome;
- d) comparing said test signature with a reference signature of a healthy subject and/or a subject suffering from the digestive tract or gastrointestinal
30 disorder, or systemic disease and classifying the test signature as a

signature of a healthy subject or as a signature of a subject suffering from a digestive tract or gastrointestinal disorder, or systemic disease.

Embodiment 30. The method of embodiment 29, wherein said digestive tract
 5 or gastrointestinal disorder, or systemic disease is selected from Crohn's disease, colitis ulcerosa, collagenous colitis, lymphocytic colitis, ischemic colitis, microscopic colitis, Behçet's syndrome, diversion colitis, diverticular colitis, eosinophilic colitis, radiation colitis, diverticulitis, irritable bowel syndrome (IBS), coeliac, lactose intolerance, Necrotising Enterocolitis
 10 (NEC), Clostridium Difficile Associated Diarrhea, colorectal cancer, attention-deficit/hyperactivity disorder (ADHD), obsessive-compulsive disorder (OCD), anxiety, stress, eating disorder, major depressive disorder (MDD), bipolar disorder, depression, schizophrenia, arthritis, sarcoidosis, mixed connective tissue disease, spondylitis ankylopoetica, osteoporosis,
 15 juvenile idiopathic arthritis, osteoarthritis, rheumatoid arthritis, sepsis, Behçet's disease, Sjögren's syndrome, fibromyalgia, sclerodermia, Raynaud's phenomenon, and systemic lupus erythematosus.

Embodiment 31. A method of typing a sample of the skin or wound flora of a
 20 subject for having the signature of a skin disorder or chronic wound, said method comprising analyzing the composition of the population of microorganisms in the intestinal microbiome based on taxonomic variation in the DNA sequence of the microbial 16S-23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of said microorganisms, wherein
 25 the sequences of conserved DNA regions comprised in the 16S and 23S rRNA sequences flanking said ITS region in the microbial genome comprise primer binding sites for amplification of said ITS regions, said analysis comprising the steps of:
 a) providing a sample of genomic DNA from the population of
 30 microorganisms constituting the intestinal microbiome of a subject;

- b) performing a PCR amplification reaction on said sample of genomic DNA using at least one set of PCR amplification primers directed to said flanking conserved DNA regions to thereby amplify and provide amplification products of said ITS regions comprised in said genomic DNA sample;
- 5 c) analyzing said amplification products to thereby provide a test signature of the composition of the population of microorganisms in said intestinal microbiome;
- d) comparing said test signature with a reference signature of a healthy subject and/or a subject suffering from a skin disorder or chronic wound and
- 10 classifying the test signature as a signature of a healthy subject or as a signature of a subject suffering from a skin disorder or chronic wound preferably wherein said skin disorder is psoriasis, eczema, acne, or rosacea, or wherein said chronic wound is a diabetic ulcer, or ulcer associated with vascular disease.

15

- Embodiment 32. A method of typing a sample of a subject for determining fertility or the probability of success of an *in vitro* fertilization or embryo transfer procedure in a female, for diagnosing or monitoring bacterial vaginosis, for diagnosing or monitoring an oral, nasal or oropharyngeal
- 20 disorder selected from periodontitis, periimplantitis, and oronasopharyngeal carcinoma for diagnosing or monitoring a disorder that affects the upper or lower respiratory tract, selected from respiratory infection, chronic obstructive pulmonary disease (COPD), asthma, cystic fibrosis and lung cancer,
- 25 said method comprising analyzing the composition of the population of microorganisms in said sample obtained from said subject based on taxonomic variation in the DNA sequence of the microbial 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of said microorganisms, wherein the sequences of conserved DNA regions
- 30 comprised in the 16S and 23S rRNA sequences flanking said ITS region in

the microbial genome comprise primer binding sites for amplification of said ITS regions, said analysis comprising the steps of:

- a) providing a sample of genomic DNA from the population of microorganisms constituting the intestinal microbiome of a subject;
- 5 b) performing a PCR amplification reaction on said sample of genomic DNA using at least one set of PCR amplification primers directed to said flanking conserved DNA regions to thereby amplify and provide amplification products of said ITS regions comprised in said genomic DNA sample;
- c) analyzing said amplification products to thereby provide a test signature
- 10 of the composition of the population of microorganisms in said intestinal microbiome;
- d) comparing said test signature with a reference signature of a beneficial sample having the desired signature and classifying the test signature.

15 Embodiment 33. The method of any one of embodiments 25-32, wherein said step b) of performing a PCR amplification reaction on said sample of genomic DNA using at least one set of PCR amplification primers directed to said flanking conserved DNA regions comprises the use of a set of PCR amplification primers for amplifying the 16S–23S rRNA internal

20 transcribed spacer (ITS) regions in the genomic DNA of the phyla *Firmicutes* and *Bacteroidetes* and a set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phylum *Proteobacteria*.

25 Embodiment 34. The method of embodiment 33, wherein said set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phyla *Firmicutes* and *Bacteroidetes* comprises the use of:

- a) the forward primer 5'-CTGGATCACCTCCTTTCTAWG-3'
- 30 comprising a first fluorescent label,

b) the forward primer 5'-CTGGAACACCTCCTTTCTGGA-3' comprising a second fluorescent label;

c) and three unlabeled reverse primers 5'-AGGCATCCACCGTGCGCCCT-3'; 5'-AGGCATTCACCRGTGCGCCCT-3'; and
 5 5'-AGGCATCCRCCATGCGCCCT-3'.

Embodiment 35. The method of embodiment 33 or 34, wherein said set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phylum

10 *Proteobacteria* comprises the use of:

a) the forward primer 5'-CCGCCCCGTCACACCATGG-3'

b) at least one of the reverse primers selected from the group consisting of

5'-AATCTCGGTTGATTTCTTTTCCT-3',

5'-AATCTCGGTTGATTTCTTCTCCT-3',

15 5'-AATCTCTTTTGATTTCTTTTCCTCG-3',

5'-AATCTCATTTGATGTCTTTTCCTCG-3',

5'-AATCTCTTTTGATTTCTTTTCCTTCG-3',

5'-AATCTCTCTTGATTTCTTTTCCTTCG-3',

5'-AATCTCAATTGATTTCTTTTCCTAAGG-3',

20 wherein at least one of said primers comprises a fluorescent label, preferably wherein said fluorescent label differs from fluorescent labels of other phylum-specific primers in said reaction.

Embodiment 36. The method of any one of embodiments 25-35, wherein the
 25 signature of a subject suffering from diverticulitis is characterized in having an increased diversity in the phylum *Proteobacteria* as calculated using the Shannon index.

Embodiment 37. The method according to embodiment 36, wherein the increased diversity in the phylum *Proteobacteria* is due to an increased diversity in the family *Enterobacteriaceae*.

- 5 Embodiment 38. The method of embodiment 36, wherein an increased diversity in the phylum *Proteobacteria* is indicated by an increase in the presence of at least one of the species selected from the group consisting of *Escherichia coli*, *Klebsiella pneumoniae*, *Enterobacter aerogenes*, *Serratia marcescens*, *Klebsiella variicola*, *Providencia stuartii*, *Desulfovibrio* spp.,
 10 *Stenotrophomonas* spp. (*Xanthomonas* spp.), *Pseudomonas aeruginosa*, *Burkholderia* spp. and *Aggregatibacter actinomycetemcomitans*.

- Embodiment 39. The method of any one of embodiments 25-38, comprising the use of a PCR calibrator system as defined in embodiment 1 or 2 in
 15 accordance with a method as described in any one of embodiments 4-9.

- Embodiment 40. The method of any one of embodiments 25-39, wherein step a) of providing a sample of genomic DNA from the population of microorganisms constituting the intestinal flora of a subject comprises the
 20 sampling of the intestinal flora using a rectal swab.

- Embodiment 41. A method of sampling the intestinal flora for analyzing the taxonomic composition of microorganisms microbiome based on taxonomic variation in the DNA sequence of the microbial 16S-23S rRNA internal
 25 transcribed spacer (ITS) regions in the genomic DNA of said microorganisms, wherein said sample of intestinal flora is obtained by using a rectal swab.

- Embodiment 42. A set of oligonucleotide primers for amplifying sequences of
 30 the 16S-23S intergenic region of the phylum *Proteobacteria* from a sample of

microbial genomic DNA, said set comprising the forward primer having the sequence 5'-CCGCCCCGTCACACCATGG-3' and at least 2 reverse primers selected from the group of primers indicated having the sequence

- 5'-AATCTCGGTTGATTTCTTTTCCT-3',
 5 5'-AATCTCGGTTGATTTCTTCTCCT-3',
 5'-AATCTCTTTTGATTTCTTTTCCTCG-3',
 5'-AATCTCATTTGATGTCTTTTCCTCG-3',
 5'-AATCTCTTTTGATTTCTTTTCCTTCG-3',
 5'-AATCTCTCTTGATTTCTTTTCCTTCG-3',
 10 5'-AATCTCAATTGATTTCTTTTCCTAAGG-3',

wherein preferably at least one of said primers comprises a detectable label.

- For the purpose of clarity and a concise description features are described herein as part of the same or separate embodiments, however, it
 15 will be appreciated that the scope of the invention may include embodiments having combinations of all or some of the features described herein. For instance, the primers and probes as described herein, such as the phylum-specific *Proteobacteria* primers, may optionally be used in any of the aspects described herein. Likewise, the calibrator system is envisioned
 20 as a preferred embodiment in any of the methods and products described.

- The invention will now be illustrated by the following example, which is provided by way of illustration and not of limitation and it will be understood that many variations in the methods described and the amounts indicated can be made without departing from the spirit of the invention
 25 and the scope of the appended claims.

EXAMPLES

Example 1. PCR calibrator system

This example describes in more detail an embodiment of the PCR calibrator system of the invention.

Single copies of 5 PCR calibrators varying in length from 190, 300, 480, 630, and 820 nucleotides, each calibrator including adaptor sequences for binding of amplification primers, was inserted into a plasmid, pUC18. The PCR calibrators were based on one of the 16S-23S ITS sequences of the bacteria *Bacillus cereus*, *Staphylococcus epidermidis*, *Parabacteroides distasonis*, *Veillonella parvula*, *Staphylococcus lugdunensis* and *Escherichia coli*. The adaptor sequences consisted of sequences complementary to the primers MPICf (5'-GACCTAGTGGAGGAAAGATAC-3'; forward primer, labeled with a fluorescent label) and MPICr (5'-TCCGTAGGTGGCACGCGGGA-3'; reverse primer, unlabeled).

The individual PCR calibrator sequences with flanking adaptors were divided in the plasmid by restriction sites for the restriction enzymes Not1, Nde1 and Sca1. The plasmid was cloned and replicated in *E. coli* cells of the strain DH5 α . Plasmid DNA was then isolated and the individual PCR calibrator fragments were released by digestion with the restriction enzyme. This resulted in a mixture of 5 PCR calibrators, each having different length and present in exact equal amounts.

The PCR calibrators were amplified using the primers MPICf and MPICr described above in the presence of a representative (inhibiting) genomic DNA sample (genomic DNA from intestinal flora), and in a non-inhibiting control environment (demineralized water). Amplifications were carried out on a GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA). Cycling conditions for PCR were 72°C for 2 min; 35 cycles of 94°C for 30 s, 56°C for 45 s, and 72°C for 1 min; and a final extension at 72°C for 5 min. The PCR mixture, with a final volume of 25 μ l, contained 10 μ l of buffered DNA (test) or demineralized water (control), 1x superTaq buffer (SphaeroQ, Gorinchem, the Netherlands), 200 μ M deoxynucleoside

triphosphate, 0.04% BSA, 1 U of superTaq, and 0.13 μ M of each of the two primers.

Following amplification, and optional purification of the amplification products, 5 μ l of the PCR product was mixed separately with
5 19.9 μ l formamide and 0.1 μ l Mapmaker 1000 ROX labeled size marker (BioVentures, Murfreesboro, TN, USA). Subsequently, DNA fragment analysis was performed on both samples with an ABI Prism 3130XL Genetic Analyzer (Applied Biosystems). All data were pre-processed with the IS-pro proprietary software suite (IS-Diagnostics, Amsterdam, the Netherlands).
10 This process resulted in profiles consisting of peaks with a specific length, measured in nucleotides, reflecting lengths of IS fragments, and a specific height, measured in relative fluorescence units (RFU), reflecting quantity of PCR product.

15 Figure 1A shows the profile as obtained for the test sample containing the five PCR calibrator fragments after amplification and length sorting by capillary gel electrophoresis. It is clear that fragments of different length were amplified with different efficiency and yielded peaks of differing height.

20 It is shown in Figure 1B how this profile is used to provide a multi-linear model for correction of the amplification efficiency of fragments of all sizes. It is shown in Figure 1C how this profile is used to provide a non-linear model for correction of the amplification efficiency of fragments of all sizes. It is considered that the non-linear model for correction of
25 amplification efficiency leads to more accurate estimations of amplification efficiency. It is further shown in Figure 1D how a corrective factor as estimated by the amplification function is applied to all peaks in an IS-pro profile. This corrective factor is applied to each peak, and is equivalent to the distance between the amplification efficiency function and the highest
30 point of that function for the fragment length of the peak to be corrected

(left panel). When applied to a profile using equal amounts of calibrator, all PCR calibrator fragments will be equal in height. Other fragments will gain a height correction dependent on their position on the x-axis, corresponding with their fragment length (Figure 1E).

5 The use of the PCR calibrator system for upscaling of partially inhibited samples is explained in Figure 2. In panel A, a blank (uninhibited) sample is displayed, while in panel B, a test (inhibited) sample is displayed. When the height of PCR calibrator peaks in a (partially) inhibited sample are compared to those in a blank sample, a correction factor can be applied
10 over the entire range of fragment length towards the height of the PCR calibrator peaks as observed in the uninhibited (blank) sample.

 The use of the PCR calibrator system for error checking of automated length marker detection (by pre-processing software) is explained in Figure 3. Figure 3A shows the raw data for the length markers
15 following capillary gel electrophoresis. The red peak pattern is the fluorescent signal of individual markers where the x-axis represents the transition time across the capillary matrix. The intensity of the signal is displayed on the Y-axis, and represents the amount of fluorescent product. Figure 3B is a graphic presentation of the profile after pre-processing using
20 dedicated software to identify individual peaks and assign to each peak its respective length. Figure 3C shows a graphic presentation of the same profile wherein the fluorescence channel of the PCR calibrator peaks is included. The PCR calibrators have been run in the capillary together with the length marker. The software assigns lengths to the PCR calibrator
25 peaks based on its detection of the length marker. These assigned PCR calibrator lengths can then be compared to the known (true) lengths of the PCR calibrator fragments. If the assigned and true lengths are the same, the software has preformed the length marker detection correctly.

Example 2. Typing diverticulitis in a subject using PCR-based profiling.

This experiment was part of a multicenter randomized clinical trial investigating the cost-effectiveness of treatment strategies with or without antibiotics for uncomplicated acute diverticulitis. We carried out this prospective cohort study in three centers (one academic hospital and two teaching hospitals).

Subjects

Eligible diverticulitis patients were consecutive trial subjects from the three including centers of 18 years or older with a first episode of acute left-sided uncomplicated modified Hinchey 1A or 1B (Wasvary H, Turfah F, Kadro O, et al. 1999. Am Surg 65:632-5) diverticulitis demonstrated by computed tomography (CT). Informed consent was obtained from these trial subjects.

The control subjects were derived from an existing database of a mixed population of adult patients evaluated in another academic hospital for a range of gastrointestinal complaints, notably with no diagnosis of diverticulitis. Diverticulosis is a common finding at colonoscopy, with a prevalence of DD that increases with age from less than 10% in people younger than 40 years to 50-66% in octogenarians. The lifetime risk to develop diverticulitis is less than 25% in these patients. Possibly a continuum in the microbiota composition exists in patients with diverticulosis and diverticulitis. To incorporate the possibility to distinguish mild diverticulitis from diverticulosis the control group also included patients with diverticulosis.

The indications for and/or the diagnoses after colonoscopy in the control subjects were the following: follow-up after polypectomy ($n=1$), anaemia e.c.i. ($n=1$), benign neoplasm ($n=3$), malignant neoplasm ($n=1$), Morbus Crohn ($n=4$), ulcerative colitis ($n=2$), indeterminate colitis ($n=1$),

irritable bowel syndrome ($n=2$), abdominal pain e.c.i. ($n=1$), surveillance for familial cancer susceptibility ($n=3$) and diverticulosis ($n=6$).

Rectal swabs

5 In the diverticulitis patients, sampling of the intestinal flora by means of a rectal swab was performed by using standard medical methods immediately at presentation on the emergency ward, and prior to starting antibiotics when allocated to this treatment. The control subjects had their rectal swab taken prior to colonoscopy which was performed to evaluate
10 their gastrointestinal complaints or for other indications. Swabs were gathered in sterile containers with 1 ml of reduced transport fluid (RTF) medium (Syed SA, Loesche WJ. 1972. Appl Microbiol 24:638-44) and stored at -20 °C within two hours of collection.

15 *DNA isolation and PCR amplification*

After thawing of the samples, total DNA isolation was carried out on all samples by an automated isolation procedure (EasyMag, Biomérieux/MP96, Roche). The principle behind the isolation includes lysis with guanidine isothiocyanate based buffers followed by automated DNA extraction with
20 magnetic beads. The DNA was eluted in 110 µl buffer and stored at 4 °C until use for polymerase chain reaction (PCR) amplification. For all samples the procedure as described by Budding et al. 2010 was used. This procedure consisted of two multiplex PCRs: a first PCR for the phyla *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Fusobacteria* and *Verrucomicrobia* was
25 performed with 10 µl of eluted DNA. Two labeled forward primers and three universal unlabeled reverse primers were used for the amplification of the interspacer (IS) region between 16S and 23S. The labeled primers were specific for the phylum *Bacteroidetes* and for the combined cluster of phyla *Firmicutes/Actinobacteria/Fusobacteria/Verrrucomicrobia*, respectively. A
30 second, separate PCR was performed for the phylum *Proteobacteria*. For this

reaction, 5 µl of eluted genomic DNA was added to the same PCR mix as used for the first PCR, only with differing primers: one labeled forward primer and six reverse primers for amplification of the IS region between 16S and 23S in *Proteobacteria*, as described herein above. After
5 amplification, 5 µl of each PCR product was mixed separately with 19.9 µl formamide and 0.1 µl Mapmaker 1000 ROX labelled size marker (BioVentures, Murfreesboro, TN, USA). Subsequently, DNA fragment analysis was performed on both samples with an ABI Prism 3130XL Genetic Analyzer (Applied Biosystems).

10

Data analysis

Log2 transformation and phylum abundance

All data were pre-processed with the IS-pro proprietary software suite (IS-Diagnostics, Amsterdam, the Netherlands). This process resulted
15 in profiles consisting of a set of 1071 peaks with a specific length, measured in nucleotides, reflecting lengths of IS fragments, and a specific height, measured in relative fluorescence units (RFU), reflecting quantity of PCR product. In order to further analyze the data, we considered each peak in a profile as an operational taxonomic unit (OTU) and its corresponding
20 intensity as its abundance. All intensities were log2 transformed. Log2 transformation of complex profiles compacts the range of variation in peak heights, reducing the dominance of high peaks and including less abundant species of the microbiota in downstream analyses. This results in improved consistency of estimated correlation coefficient, lower impact of inter-run
25 variation and improved detection of less prominent species. This conversion was used in all downstream analyses such as calculating within-sample and between-sample microbial diversity. The clustered heat map was made by generating a correlation matrix of all log2 transformed profile data followed by clustering with the unweighted pair group method with arithmetic mean
30 (UPGMA).

Diversity Analysis

Diversity was calculated both per phylum and per the overall microbial composition (by pooling all phyla together). Within-sample
 5 diversity was calculated as the Shannon index, that was recently shown to be a robust estimate of microbial diversity (Haegeman B, Hamelin J, Moriarty J, et al. 2013. ISME J 7:1092-101). Dissimilarities between samples, or between-sample diversity, was represented in a dissimilarity matrix that was built using the cosine distance measure. Given two vectors
 10 of attributes (two profiles in our case), A and B, the cosine dissimilarity is represented using a dotproduct and magnitude as:

$$\text{dissimilarity} = 1 - \cos(\theta) = 1 - \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

The resulting dissimilarity matrix was summarized and visualized in a low-dimensional space using principal coordinate analysis (PCoA). Diversity analysis was performed using the vegan software package in R.

15

Partial least squares discriminant analysis (PLS-DA)

A partial least squares discriminant analysis (PLS-DA) regression model was used for the prediction of clinical status of samples; i.e. whether it belonged to a diverticulitis patient or to a control subject. PLS-DA is a
 20 supervised classification method, which aims to find linear transformations of a matrix of predictors and categorical responses so as to maximize their covariance. [Rajilić-Stojanović M, Biagi E, Heilig HG, et al. 2011. Gastroenterology 141:1792-801] The PLS-DA model was constructed on the basis of four different datasets: one for each of the three separate phylum
 25 groups and one for the overall microbial composition, by pooling all phyla. Under the assumption that the more discriminant variables are the ones with a higher variance, we performed a preliminary variable selection by

filtering out low variance predictors. Only the top 25% most variable predictors were considered in the analysis.

PLS-DA model validation was carried out by a 10-fold cross validation procedure. In practice, the dataset was split into 90% of samples for model construction (i.e. the training set) with the aim to predict the other 10% (i.e. the test set). This procedure was repeated for 10 iterations, where each sample served as a test sample exactly once. Accuracy rates, specificity and sensitivity were computed for the samples that were used as a test set in every iteration, and the model predictive power was further assessed using a receiver operating characteristic (ROC) curve, a function of the true positive rate (TPR or sensitivity) and false positive rate (FPR or 1-specificity).

PLS-DA provides a quantitative estimate of the discriminatory power of each descriptor by means of VIP (variable importance for the projection) parameters. VIP values rank the descriptors by their ability to discriminate different groups and are therefore considered an appropriate quantitative statistical parameter. We used the VIP criterion to rank the different OTUs based on their contribution to the response variable (clinical status, i.e. diverticulitis: yes or no) and PLS components. Only the OTUs with the highest contribution (VIP score > 1.2) were considered. The OTUs resulting from this selection were translated to most likely bacterial species by comparison to a database consisting of >1500 bacterial species and their associated IS lengths. PLS-DA analysis was performed using the DiscrMiner package in R (version 2.15.2). All data visualizations were performed with the Spotfire software package (TIBCO, Palo Alto, CA, USA).

Results

Bacterial phylum abundance and profile clustering

The *Firmicutes* to *Bacteroidetes* ratio is commonly used to describe and characterize a dysbiosis of the gut microbiota in different

disease states, such as irritable bowel syndrome (IBS) and obesity (Ley et al. 2006. Nature 444:21-28; Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. New York: Springer- Verlag 2001]. Since these two phyla are being amplified in the same PCR reaction, we could compare their relative abundance between patients and controls. The phylogenetic characterization of samples from control subjects uncovered that *Bacteroidetes* represented 51% and *Firmicutes* 49% of the total abundance in the *Firmicutes/Bacteroidetes* PCR. Exactly the same proportions were found for the patient group. The total load of bacteria of the *Proteobacteria* phylum was relatively similar between patients and controls ($10.2 \pm 1.9 \log_2$ RFU and $10.1 \pm 2.0 \log_2$ RFU for patients and controls, respectively; $p=0.20$, Mann-Whitney U-test).

A heat map was generated from all IS-profiles separated by phylum. IS-profiles showed a general separation of samples from diverticulitis patients and controls when clustering was performed by *Proteobacteria* profile (Figure 4).

Microbial diversity and composition in diverticulitis patients versus controls

While diversity of the phyla *Bacteroidetes* or *Firmicutes* did not differ between patients and controls, the Shannon index indicated that the diversity of the *Proteobacteria* phylum was significantly higher in patients compared to controls (2.6 (IQR 1.07) and 3.2 (IQR 0.5) for controls and patients respectively; $p<0.00002$, Mann-Whitney U-test), which also affected the difference in diversity measured when considering all phyla together (3.9 (IQR 0.3) and 4.1 (IQR 0.3) for controls and patients respectively; $p<0.002$, Mann-Whitney U-test) (Figure 5).

Differences in overall bacterial community composition were assessed using cosine distances between samples. PCoA did not segregate diverticulitis patients and controls into different groups for the phyla *Bacteroidetes* and *Firmicutes*. However, patients could be clustered

separately from controls in a 3-dimensional space based on their *Proteobacteria* profiles (Figure 6).

Discriminative ability of PLS-DA

5 The use of an unsupervised approach for classification (PCoA) already demonstrated diagnostic potential of *Proteobacteria* profiles in predicting the health status of a given patient. This potential was born out in a supervised analysis, using PLS-DA known to be suitable for high-dimensional data (Rajilić-Stojanović et al. 2011. Gastroenterology 141:1792-10 801; Pérez-Enciso et al. 2003. Hum Genet 112:581-92; Nguyen et al. 2002. Bioinformatics 18:39-50). The PLS-DA model used 268 OTUs, representing the 25% most variable OTUs, as predictors and the clinical status of the samples (i.e. diverticulitis: yes or no) as the response variable. In order to quantify the discriminative ability of the model we first considered the full 15 datasets (three individual phylum datasets, and one composed of all phyla). Taking the *Bacteroidetes* or *Firmicutes* data as input resulted in low predictive accuracy rates (55% and 53% for *Bacteroidetes* and *Firmicutes*, respectively; data not shown). Taking the *Proteobacteria* data as input resulted in a predictive accuracy rate of 95% (Figure 7). Three out of 56 20 samples were misclassified: one control and two patients, whose samples are the encircled ones in the PCoA scatterplot (Figure 6). Resulting specificity was thus calculated to be 96% with a sensitivity of 94%. Taking the combined dataset, composed of all three phyla, as input resulted in an accuracy rate of 96% with two misclassified controls, corresponding to a 25 specificity of 92% and 100% sensitivity. The misclassified controls were two subjects with diverticulosis. The most discriminative OTUs were found to derive largely from the family *Enterobacteriaceae* (Table 1).

Table 1. Most discriminative OTUs based on a Variable Importance for
30 Projection value >1.2

Species	Family
<i>E.coli</i>	<i>Enterobacteriaceae</i>
<i>K.pneumoniae</i>	<i>Enterobacteriaceae</i>
<i>Enterobacter aerogenes</i>	<i>Enterobacteriaceae</i>
<i>S.marcescens</i>	<i>Enterobacteriaceae</i>
<i>Klebsiella variicola</i>	<i>Enterobacteriaceae</i>
<i>Providencia stuartii</i>	<i>Enterobacteriaceae</i>
<i>Desulfovibrio sp.</i>	<i>Desulfovibrionaceae</i>
<i>Xanthomonas</i>	<i>Xanthomonadaceae</i>
<i>Stenotrophomonas</i>	<i>Xanthomonadaceae</i>
<i>Pseudomonas aeruginosa</i>	<i>Pseudomonadaceae</i>
<i>Burkholderia sp.</i>	<i>Burkholderiaceae</i>
<i>Aggregatibacter actinomycetemcomitans</i>	<i>Pasteurellaceae</i>
Unknown <i>Proteobacteria</i> species*	Unknown

*11 types of unknown *Proteobacteria* species were identified.

Prediction of diverticulitis using the PLS-DA model

The predictive ability of the model was assessed by cross-validation. The dataset was randomly divided into 10 subsets, such that in each iteration one subset was chosen to serve as a test set. The prediction results from all iterations were then pooled together and enabled us to estimate the performance of the model by means of predictive power. Taking account of the *Bacteroidetes* or the *Firmicutes* phylum only resulted in a low predictive accuracy. For both the *Bacteroidetes* and the *Firmicutes* the cross-validated accuracy rate was 51%. Considering only the *Proteobacteria*, we reached a cross-validated accuracy rate of 80%. Six controls and five patients were wrongly classified, which resulted in a specificity of 76% and a sensitivity of 84%. When we combined the three phyla, we could reach a diagnostic accuracy rate of 84% (spec=80%; sens=87%). Figure 8 summarizes the predictive power of the PLS-DA model by means of ROC curves.

The results of this study indicate that the fecal microbiota diversity of patients with a first episode of acute uncomplicated left-sided diverticulitis differs significantly from control subjects from a general

gastroenterological practice, with the *Proteobacteria* phylum defining this difference. Furthermore, predictive diagnostics based on complex microbiota data seems feasible for diagnosing diverticulitis, with a diagnostic accuracy rate of 84%. The most discriminative species derived from the family

5 *Enterobacteriaceae*.

Several studies have identified characteristics of the intestinal microbiota that may be associated with disease, but clinical diagnostic tools based on microbiome analysis still need to be developed. Whereas most studies into microbiota composition in health and disease identified

10 correlations, here we demonstrate an approach in which microbiota composition may be used as a clinical predictor. By employing a supervised algorithm in combination with cross-validation, we show how microbiota analysis may move towards prediction instead of correlation. PLS regression provides a dimension reduction strategy in situations where a set of

15 response variables need to be related to a set of predictor variables. It is considered a supervised learning method since it uses the dependent (clinical status in this study) as well as the independent variables (OTUs) to construct variable selection and importance ranking. PLS-DA refers to the particular case where the response variable is a set of binary variables

20 describing the categories of a categorical variable, e.g. disease states. This model is commonly used in the field of chemometrics and in the analysis of microarray expression data, as it is especially suited to deal with a much larger number of predictors than observations and with multi-collinearity. In this study we encountered similar challenges; the number of OTUs is much

25 larger than the number of samples and some of them are highly correlated. Due to the properties mentioned above, we found this approach also very appropriate to apply to IS-pro data. The VIP criterion was previously used in PLS-DA microarray analyses to assess which genes were useful to discriminate between different groups.

Specific shifts in the phylum *Proteobacteria* -other than general measures like diversity- have not been found to be associated with disease before. This might be caused by the fact that *Proteobacteria* generally have a low relative abundance in the intestinal microbiota. Because almost all

5 current approaches to analyze the intestinal microbiota use universal bacterial amplification as a starting point, low abundant phyla such as the *Proteobacteria* remain relatively under explored as other, more prevalent taxa will dominate the PCR reaction and following analyses. In contrast, the IS-pro molecular technique comprises two separate phylum-specific PCR

10 reactions: one for the amplification of *Bacteroidetes/Firmicutes* and another for the specific amplification of *Proteobacteria*. While the separation of the different phyla in two PCRs prevents us from addressing all three phyla together when presenting their relative abundances -consequently hampering direct comparisons of abundances- it does allow us to zoom in

15 and analyze the *Proteobacterial* community composition in-depth.

Brook *et al.* retrospectively studied the aerobic and anaerobic microbiology of 110 specimens from the peritoneal cavity after intestinal perforation and in 22 specimens from abdominal abscesses of patients with complicated diverticulitis [Brook I, Frazier EH. 2000. J Med Microbiol

20 49:827–30] With conventional culture techniques they identified *E. coli* and *Streptococcus spp* as the predominant aerobic and facultative bacteria and the most frequently isolated anaerobes were *Bacteroides spp* (*B. fragilis* group), *Peptostreptococcus*, *Clostridium* and *Fusobacterium spp*. The only study up to date with PCR based sequencing of the microbiota in

25 diverticulitis patients was conducted by Gueimonde *et al.* [Gueimonde M, Ouwehand A, Huhtinen H, et al. 2007. World J Gastroenterol 13:3985-9] They identified a significant higher occurrence of *Bifidobacterium longum* and *Bifidobacterium animalis* in patients with diverticulitis, and their overall conclusion was that aberrancies in mucosa associated microbiota are

30 present in different intestinal diseases. However, in their study only nine

diverticulitis patients were included. Resected mucosal samples were compared with those of 21 colon cancer patients and four inflammatory bowel disease patients, but no healthy controls. Surprisingly, they looked only at the genus *Bifidobacterium* and did not analyze the entire profile; they stated they used the bifidobacterial microbiota as an indicator of alterations in the mucosal colonisation pattern. The bifidobacterial microbiota however, is known to constitute only a small fraction of the intestinal microbial composition in adults.

Currently, antibiotics are often used in the conservative treatment of uncomplicated diverticulitis despite the lack of sound evidence. Cyclic administration of rifaximin has been proven to be effective in reducing symptoms and complications and possibly prevents recurrence in patients after complicated diverticulitis. Relatively new therapies, such as probiotic therapy, are proposed as well for the management of diverticular disease (DD). Indeed, a few small open label studies already show promising results. Considering that antibiotic and probiotic treatments are regularly prescribed to DD patients, it is striking that relatively few studies have been performed to improve our understanding of the composition of the colonic microbiota. The pathophysiology of diverticulitis was assumed to be clear and well understood but actually astonishingly little is known about causal factors for this disease. Our understanding of the effect of changes in microbiota abundance, diversity and composition is limited. Our study therefore, is a first step in further elucidating the etiopathogenesis of diverticular disease and its inflammatory complications.

Since a clinical diagnosis of diverticulitis can not be made with a high certainty without imaging, it seems appropriate to evaluate a test intended for making a specific clinical diagnosis against a patient group with variable clinical presentation. By taking a cross-section of patients in a general gastroenterological practice instead of a healthy control group, the specificity of the prediction becomes more meaningful.

This study has some limitations. First, we have data on only a small study group. As a result we are not able to estimate and optimize predictive ability robustly. The performance of a predictive tool is prone to be overestimated in its own study cohort. For diagnostics by microbiome to be applied in daily practice a study like this one should be externally validated and followed by a larger study to confirm results and calculate sensitivity and specificity more robustly. Second, as a consequence of a small sample size, we were not able to firmly compare diverticulitis patients with subjects with diverticulosis. It has been hypothesized that DD patients have a changed colonic microbiome. From an etiopathogenetic point of view, it would be informative to know to what extent the microbiome in diverticulosis resembles the microbiome in diverticulitis or health. Indeed, the two controls that were misclassified, were subjects with diverticulosis. This seems to underline a shift in microbiota related to diverticular disease. The present tool may reveal a gradual shift in microbiota composition from patients with diverticulosis towards diverticulitis. In the present study, species identification was done by *in-silico* comparison of fragment lengths. This technique generally gives consistent results.

The present study demonstrates that the diagnosis of diverticulitis can be done by microbiome analysis with relatively good accuracy. More generally, this study illustrates a proof of concept of how diagnostics based on complex microbiota data in a broader sense may be applied. Thus, the fecal microbiota can be used as diagnostic tool for diverticulitis, with patient stratification directing a personalized treatment strategy, whether or not to prescribe antibiotics, the type of antibiotic, and even to monitor disease course. Clinical application as a diagnostic tool also reduces the need for imaging to diagnose diverticulitis.

Example 3. A convenient method for reproducible sampling of the intestinal microbiota in a clinical setting.

This study was set up as a descriptive study to compare rectal swabs for sampling of the intestinal microbiota with sampling of feces specimens and mucosal biopsies. Two rounds of investigations were held. In the first, patients who underwent an elective colonoscopy were asked to
5 bring in feces one week before the procedure, and rectal swabs and biopsies were taken during the procedure. Because colonoscopy involves extensive intestinal preparation, in a second round patients with inflammatory bowel disease were asked to sample feces and to obtain one rectal swab at home; a second rectal swab was obtained on the day they brought in the feces
10 (Figure 9).

Subjects were included that either underwent an elective colonoscopy between February and June 2011 (for sets of rectal swabs, mucosal biopsies and fecal samples in prepped patients) or presented for commonly scheduled control at the inflammatory bowel disease (IBD)
15 outpatient clinic in October 2012 (for sets of rectal swabs and feces in unprepped patients). In the first group, the only exclusion criterion was a contraindication for taking mucosal biopsies.

Samples

20 Rectal swabs were taken with FLOQSwabs 552C (Copan, CA, USA), which were inserted into the anal canal, beyond the anal verge (± 3 cm). All patients who underwent colonoscopy were prepped according to a standardized protocol with a laxative preparation consisting of high-volume polyethylene glycol (PEG) solution (Kleanprep®, Helsinn, Lugano,
25 Switzerland). Mucosal biopsy specimens were collected with a flexible video endoscope (Olympus GmbH, Hamburg, Germany) and a flexible biopsy forceps (Wilson-Cook; European Endoscopy Group, Fujinon Medical Holland, Veenendaal, The Netherlands). Mucosal biopsy specimens were harvested from sigmoid colon at 20-30 cm from the anal verge. Per subject,
30 one mucosal sample was washed twice in 500 μ l PBS (pH 7) before snap-

freezing in liquid nitrogen and a second sample was deposited in a container filled with 500 µl PBS and snap frozen in liquid nitrogen. All samples were stored at -20°C . In the colonoscopy group, rectal swabs were collected at the time of colonoscopy, just prior to the endoscopic procedure. Two rectal swabs
5 were deposited in a container with 500 µl Reduced Transport Fluid (RTF) buffer⁹ and kept at room temperature for 2 hours prior to storage at -20°C . One swab was immediately snap frozen in liquid nitrogen (with no RTF buffer).

In the IBD outpatient group, two rectal swabs were gathered. One
10 was taken by the patients themselves, at home, one day prior to presentation and stored in RTF buffer at -20°C . The other was taken at the outpatient clinic and stored in the same fashion. Both swabs were stored in a container with 500 µl RTF buffer at -20°C .

In both the colonoscopy and the IBD outpatient group, fecal
15 samples were gathered within five days before presentation at the outpatient clinic for endoscopy or routine control. Samples were gathered in sterile containers and were stored at -20°C within 2 h after collection and kept frozen until further analysis.

20 *DNA isolation*

DNA was isolated from feces and mucosal biopsies as described in Budding et al. 2010 (cited above). In short, for mucosal samples, the first step consisted of lysis of tissue and bacteria with the QIAamp DNA mini Kit (Qiagen, Hilden, Germany) followed by DNA extraction with the
25 NucliSENS® easyMag® automated DNA isolation machine (Biomérieux, Marcy l'Etoile, France). For fecal samples, 100-400mg of feces was used as input for the fecal DNA extraction protocol of the easyMag machine as described by the manufacturer. For DNA isolation from swabs, one ml of nucliSENS® lysisbuffer, containing guanidine thiocyanate, was added to
30 each vial containing a swab tip and the mixture was shaken at 1400rpm

(Thermomixer comfort, Eppendorf, Hamburg, Germany) for five minutes. For a subset of 14 snap frozen swabs, an additional bead-beating step was evaluated. For these swabs, after five minutes of shaking at 1400 rpm, the mixture was divided into two parts. To one part, approximately 100 µg of Zirconia 0.1 silica beads were added and bead-beating was performed for 60 seconds. Afterwards, all samples were centrifuged for four minutes at 12.000g and added to the easyMag container. DNA extraction was performed on the easyMag machine with the Specific A protocol as described by the manufacturer.

IS-profiling of the intestinal microbiota

The intestinal microbiota analysis was performed by IS-pro as described in Budding et al. 2010. IS-pro involves bacterial species differentiation by the length of the 16S–23S rDNA interspace region with taxonomic classification by phylum-specific fluorescent labelling of PCR primers.

Amplification of IS regions

Five primers are used for amplification of IS regions. Two fluorescently labelled forward primers are phylum-specific, for the 16S rDNA region: one FAM-labeled primer, specific for *Firmicutes*, *Actinobacteria*, *Fusobacteria* and *Verrucomicrobia* and one HEX-labelled primer specific for *Bacteroidetes*. Three unlabeled reverse primers are specific for the 23S rDNA region. The combination of these primers provides very broad coverage for *Firmicutes*, *Actinobacteria*, and *Bacteroidetes*. The primers are used in a multiplex PCR, which amplifies the 16S-23S IS region. The length of this IS region and its PCR product is species-specific.

The fluorescent label provides identification of all fragments at the phylum level.

Amplifications were carried out on a GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA). Cycling conditions for PCR were 72°C for 2 min; 35 cycles of 94°C for 30 s, 56°C for 45 s, and 72°C for 1 min; and a final extension at 72°C for 5 min. Each PCR mixture, with a final
 5 volume of 25 µl, contained 10 µl of buffered DNA, 1x superTaq buffer (SphaeroQ, Gorinchem, the Netherlands), 200 µM deoxynucleoside triphosphate, 0.04% BSA, 1 U of superTaq, and 0.13 µM of each of the 5 primers.

10 *IS-Fragment analysis*

After PCR, 5 µl of PCR product was mixed with 19.8 µl formamide and 0.2 µl Mapmaker 1000 ROX-labeled size marker (BioVentures, Murfreesboro, TN, USA). DNA fragment analysis was performed on an ABI Prism 3130XL Genetic Analyzer (Applied Biosystems). Results are
 15 presented as color-labeled peak profiles (Figure 10). These peaks can be regarded as operational taxonomical units (OTU's) and we will refer to them as such henceforth in the text. All data were further analyzed with the Spotfire software package (TIBCO, Palo Alto, CA, USA).

20 *Statistical analysis*

All statistical analyses were performed described previously10. Comparisons between all samples (swab vs duplicate swab, swab vs snap frozen swab, swab vs feces, swab vs mucosal biopsy and feces vs mucosal biopsy) were made by calculating squared correlation coefficients for all
 25 possible pairs of samples. When duplicate swabs stored in RTF from the colonoscopy group were compared to other samples, averaged profiles of the duplicate swabs stored in RTF were used.

Comparisons were grouped in intra and inter individual comparisons, the former group comprising all comparisons between samples
 30 from the same individual, the latter group comprising all other comparisons.

Median and inter quartile range (IQR) were calculated for each comparison group.

Results

5 *Study population*

A total of 38 subjects was included in the colonoscopy group, 23 male and fifteen female. The indications for colonoscopy were suspected benign neoplasm (8), diverticulosis (6), screening for familial tumours (5), IBD (5) and general gastrointestinal complaints (7). Not all sample types
10 could be harvested from all subjects and in some samples the PCR reaction was inhibited. The total number of samples obtained was: 35 snap frozen rectal swabs, 37 sets of rectal swabs in RTF, 33 dry mucosal biopsies, 35 mucosal biopsies in sodium chloride 0.9% and 19 fecal samples.

Ten subjects were included in the IBD outpatient group, two male
15 (both with ulcerative colitis) and eight female (four with ulcerative colitis, four with Crohn's disease). Two subjects presented with active disease, eight with disease in remission. The total number of samples was: ten swabs and ten fecal samples taken at home by the patients themselves and ten swabs taken at the outpatient clinic by their physician.

20

Effect of bead-beating for DNA isolation

As it has been described that DNA isolation protocols for fecal samples that include bead-beating give higher DNA yields of certain groups of bacteria (Salonen et al. 2010 J Microbiol Methods 81:127-134), we
25 compared this procedure to automated DNA isolation for swab samples. With a mixed effects model, accounting for both fixed and random effects, we found bead-beating to have a significant negative impact on DNA recovery from *Bacteroidetes*: *Bacteroidetes* peaks in DNA isolated without a bead beating step were on average 1.85 times higher than the equivalent
30 peaks in DNA isolated with a bead beating step ($p=0.015$). For *Firmicutes*

there was also a trend towards a negative impact of bead beating.

Firmicutes peaks from DNA without bead beating were on average 1,45 times higher than equivalent peaks generated from DNA with a bead beating step ($p=0.051$) (data not shown).

5

Effect of storage and processing of rectal swabs

To evaluate the effect of a non-stringent sample storage protocol, where storage of samples at room temperature for two hours was allowed, we compared microbiota profiles of snap-frozen rectal swabs to profiles of rectal swabs that had been stored in RTF for two hours before freezing at -20°C . As storage conditions may affect Gram positive and Gram negative bacteria differently, we analyzed data separately for *Firmicutes* and *Bacteroidetes*. Correlations were calculated for profiles derived from snap frozen swab samples versus the averaged profiles of duplicate swab samples stored in RTF at room temperatures before freezing.

15

For both *Firmicutes* and *Bacteroidetes*, intra-individual comparisons showed high similarities as measured by a high median R^2 . Inter individual comparisons were low, as was expected (Figure 11B). For all measured values in the above and subsequent sections, reference is made to Table 2.

20

Comparator 1	Comparator 2	Firmicutes	Bacteroidetes
Swab	Snap frozen swab	A 0.73 {0.13}	0.33 {0.24}
		B 0.12 {0.16}	0.13 {0.20}
	Duplicate swab	A 0.70 {0.38}	0.72 {0.34}
		B 0.13 {0.15}	0.11 {0.17}
	Swab at home	A 0.55 {0.26}	0.52 {0.23}
		B 0.14 {0.05}	0.05 {0.14}
Swab at home	Faeces	A 0.17 {0.16}	0.36 {0.25}
		B 0.10 {0.12}	0.16 {0.20}
	Mucosal biopsy	A 0.15 {0.21}	0.32 {0.44}
		B 0.10 {0.11}	0.13 {0.19}
	Faeces	A 0.23 {0.12}	0.75 {0.36}
		B 0.10 {0.12}	0.05 {0.20}
Faeces	Mucosal biopsy	A 0.12 {0.28}	0.25 {0.35}
		B 0.12 {0.14}	0.18 {0.15}
	Swab at home	A 0.23 {0.12}	0.75 {0.36}
		B 0.10 {0.12}	0.05 {0.20}

Table 2: Median R squared and Inter Quartile Range (IQR) values for all comparisons. IQR values are indicated in brackets. Rows A and B indicate either intra-individual comparisons (A) or inter-individual comparisons (B).

5 *Reproducibility*

To test the reproducibility of the swabbing procedure, we compared duplicate swabs obtained from the same patients (n=37). These duplicate swabs were stored in RTF buffer at room temperature for 2 hours. Profiles were compared pairwise with Pearson correlation and results were
 10 analyzed separately for *Firmicutes* and *Bacteroidetes*. In this analysis too, intra-subject correlations were high for both *Firmicutes* and *Bacteroidetes*, whereas inter individual correlations were low (Figure 11A). These data indicate that reproducibility of rectal swabs within the same patients was very good.

15 To test whether self-sampling by patients at home would yield comparable results to swabs taken in a standardized fashion at the outpatient clinic, we compared these in ten individuals. Again, profiles were compared pairwise and results were analyzed separately for *Firmicutes* and *Bacteroidetes*. This analysis too showed high intra-subject correlations and
 20 low inter-subject correlations (Figure 11A). This showed that self-collected swabs by patients at home were highly comparable to clinically collected swabs when stored in the same fashion.

Comparison of different sample types

25 Finally, rectal swabs, fecal samples and mucosal biopsies were compared to each other. First, rectal swab microbiota profiles were compared to fecal microbiota profiles in nineteen subjects of the colonoscopy group and in all ten subjects of the IBD outpatient group. In the colonoscopy group, correlations between swab and fecal profiles were generally low.
 30 *Bacteroidetes* generally had a slightly higher correlation than *Firmicutes*

(Figure 11D). In the IBD outpatient group, all correlations were markedly higher, especially for the *Bacteroidetes*, which had a median value of R^2 similar to that found for duplicate swab profiles. As expected, inter subject correlations were low in both groups. These data show that fecal profiles resemble swab profiles, but not in people who have undergone bowel prepping (Figure 12B).

Next, we compared swab samples to mucosal biopsy samples taken by colonoscopy in 32 subjects. This comparison showed that correlation of swab profiles to mucosal profiles was low, similar to the correlation of swab samples to fecal samples in this prepped patient group. *Bacteroidetes* profiles showed a higher similarity than *Firmicutes* profiles (Figure 11C).

Finally, as we found rectal swab microbiota profiles to be distinct from mucosal and fecal microbiota profiles in the colonoscopy group, we were interested in the similarity between fecal microbiota and mucosal microbiota.

To compare these we used the same analysis as above. We found that correlations between fecal samples and mucosal biopsies were comparably low as the correlations of swab samples to both these sample types. In this comparison too, a higher correlation was found for *Bacteroidetes* than for *Firmicutes* (Figure 11E).

In this Example it is shown that rectal swabs can be conveniently taken and provided a good method to produce highly reproducible microbiota profiles which are similar to fecal sample profiles. It is therefore suggested that rectal swabs may be ideally suited for large scale studies and for routine clinical applications of microbiota profiling. Sampling of the intestinal microbiota by feces sampling or by rectal swabbing, without previous bowel preparation (without extensive bowel lavage is probably the preferred method if analyzing genuine, undisturbed microbiota in a subject (or patient). Rectal swabs are an attractive means of sampling the intestinal

microbiota in a clinical setting because they do not have the drawbacks of feces collection and harvest of mucosal biopsies. Subjects do not necessarily need to collect samples themselves as is needed for fecal samples and they do not need to be prepped or undergo invasive procedures. The applicability of rectal swabs in clinical routine settings is highlighted by the fact that rectal swabs are already commonly used in clinical routine for detection of carriage of resistant microorganisms and can be taken at every visit.

It is further shown that short-term storage of rectal swabs in RTF buffer at room temperature had no impact on the composition of the microbiota, thus relaxing requirements for sample collection and storage and making the method applicable in almost any (clinical) setting. It is demonstrated herein that storage of rectal swabs at room temperature for 2 hours in a stabilizing buffer (RTF buffer), does not impact microbiota composition.

It was found that bead beating did not contribute to the DNA yield from rectal swabs. In contrast, bead beating diminished the yield of *Bacteroidetes* DNA. This was an unexpected, yet reproducible outcome in the context of what has been described for isolation of bacterial DNA from feces. For clinical routine, in which speed and easy implementation is essential, omission of bead-beating can be considered favorable.

It is further shown that swab profiles are indeed similar to feces profiles obtained from unprepped subjects, but decidedly distinct from feces profiles in prepped patients. In these prepped subjects, microbiota profiles in swab samples were also distinct from profiles in mucosal biopsies. Rectal swabs were taken just prior to colonic mucosal biopsies, both after prepping of the subject. The difference in composition between these samples thus seemed to represent a true difference in composition between the rectum/proctum, which was sampled by rectal swabs, and more proximal in the distal sigmoid colon, as sampled by sigmoidal mucosal biopsies.

It was found that rectal swabs were a convenient means of sampling the human gut microbiota. Swabs can be taken ‘on demand’, whenever a subject presents. The acquired samples resembled fecal microbiota and showed a highly reproducible profile, whether they are gathered at home by patients or by medical professionals in an outpatient setting.

Example 4. IS-pro for analysis and diagnostics in parodontitis.

Parodontitis is an inflammation of the parodontium. It has long been known that bacteria play a causative role in this inflammatory process. In contrast to classic infection and inflammation, which is typically caused by a single causative organism, parodontitis is caused by a consortium of bacteria. The presence of the bacteria in this consortium alone may not cause harm, only in combination with other members of the consortium. The bacteria that may belong to this consortium is not a strictly defined set. Seven species have now convincingly been shown to be important players, but new organisms continue to be found that also play a role in the paropathogenic microbiota. Currently, most diagnostic tests to identify parodontitis are based on quantifying the presence of a defined number of bacteria from the paropathogenic consortium. Given the fact that the putative paropathogenic consortium has not been characterized in its entirety, it is likely that diagnostics tests that are not limited to a predefined set of microbes, may perform better in diagnosing parodontitis.

Here we evaluate the performance of the 16S-23S interspace (IS) based profiling technique IS-pro, for characterizing the population of all bacteria present in a microbial consortium in order to distinguish parodontitis from a healthy control group. We analyze data by classical statistical methods and by predictive modeling and compare results to qPCR on a defined set of seven defined members of the paropathogenic microbiota.

Materials and Methods

Design

This study was set up as a prospective cohort study, including patients with proven parodontitis and subjects with a proven healthy
5 parodontium. Two different methods were compared in their ability to distinguish parodontitis patients from the healthy controls based on parodontal microbiota.

Study population

10 Parodontitis patients and control subjects were grouped and the dental and periodontal health status of all subjects was evaluated using standard methods. Ten subjects were included in both the parodontitis and the control group.

15 *Sampling technique*

Sampling sites were isolated with cotton rolls, supragingival plaque was carefully removed with curets and the sites were air dried. Subgingival plaque samples were obtained by inserting sterile paper points into the bottom of the sulcus or pocket. After several seconds, paper points
20 were removed and deposited in a labeled sterile tube containing 1,5 ml Reduced Transport Fluid (RTF) and transported to the laboratory where they were stored at -20oC until further processing.

DNA Isolation

25 DNA was isolated from paperpoints by DNA extraction with the NucliSENS® easyMag® automated DNA isolation machine (Biomérieux, Marcy l'Etoile, France). One ml of nucliSENS® lysisbuffer, containing guanidine thiocyanate, was added to each vial containing the paperpoints and the mixture was shaken at 1400 rpm (Thermomixer comfort, Eppendorf,
30 Hamburg, Germany) by RT for five minutes. Samples were added to the

easyMag container and DNA extraction was performed on the easyMag machine with the Specific A protocol as described by the manufacturer.

Quantitative real-time PCR

5 The primer/probe sets and real-time PCR conditions were performed as described previously (Boutaga 2005). qPCR was specific for *Aggregatibacter actinomycetemcomitans*(AA), *Treponema denticola*(TD), *Fusobacterium nucleatum*(FN), *Porphyromonas gingivalis*(PG), *Parvimonas micra*(PM), *Prevotella intermedia* (PI) and *Tannerella*
10 *forsythia*(TF). In short, 3 duplex and one monoplex real-time PCR was performed with total reaction volumes of 15 µl, covering all seven species mentioned above . Duplex PCRs consisted of the combinations AA/PI, PM/PG and TF/TD. The monoplex PCR was specific for FN .

15 *IS-pro*

 Total microbiota analysis was performed by IS-pro as described herein above. IS-pro involves bacterial species differentiation by the length of the 16S–23S rDNA interspace (IS) region with taxonomic classification by phylum-specific fluorescent labeling of PCR primers. For amplification of
20 bacterial IS regions for IS-pro, two PCRs were performed. One PCR reaction amplifies IS regions of bacteria belonging to the phyla *Firmicutes*, *Actinobacteria*, *Fusobacteria*, *Verrucomicrobia* and *Bacteroidetes*. The second PCR amplifies IS regions of bacteria belonging to the highly diverse phylum *Proteobacteria*. For the first reaction, five primers were used as described
25 herein above. Two fluorescently labeled phylum-specific forward primers in the 16S rDNA region: one FAM-labeled primer, specific for *Firmicutes*, *Actinobacteria*, *Fusobacteria* and *Verrucomicrobia* (herein referred to as the *Firmicutes* phylum) and one HEX-labeled primer specific for *Bacteroidetes* (herein referred to as the *Bacteroidetes* phylum). Three unlabeled reverse
30 primers were specific for the 23S rDNA region. The combination of these

primers provides very broad coverage for the phyla mentioned above. The primers are used in a multiplex PCR, which amplifies the 16S-23S IS region. The length of this IS region and its PCR product is species-specific. The fluorescent label provides identification of all fragments at the phylum level. For the second PCR, specific for *Proteobacteria*, a single FAM labeled forward primer is used in combination with seven reverse primers as described herein above.

Amplifications were carried out on a GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA). Cycling conditions for PCR were 35 cycles of 94°C for 30 s, 56°C for 45 s, and 72°C for 1 min and a final extension at 72°C for 11 min. Each PCR mixture, with a final volume of 25 µl, contained 10 µl of buffered DNA, 1x superTaq buffer (SphaeroQ, Gorinchem, the Netherlands), 200 µM deoxynucleoside triphosphate, 0.04% BSA, 1 U of superTaq, and 0.13 µM of each of the primers.

After PCR, 5 µl of PCR product was mixed with 19.8 µl formamide and 0.2 µl Mapmaker 1500 ROX-labeled size marker (BioVentures, Murfreesboro, TN, USA). DNA fragment analysis was performed on an ABI Prism 3500 Genetic Analyzer (Applied Biosystems).

Data pre-processing

All data were pre-processed with the IS-pro proprietary software suite (IS-Diagnostics, Amsterdam, the Netherlands). This process resulted in profiles consisting of peaks with a specific length, measured in nucleotides, reflecting lengths of IS fragments, and a specific height, measured in relative fluorescence units (RFU), reflecting quantity of PCR product. In order to further analyze the data, we considered each peak in a profile as an operational taxonomic unit (OTU) and its corresponding intensity as its abundance. All intensities were log2 transformed. This conversion was used in all downstream analyses such as calculating within-sample and between-sample microbial diversity.

Heat maps and Diversity Analysis

Dendrograms of heat maps were constructed by first generating correlation matrices based on Pearson product-moment correlation coefficients (IS-pro) or Euclidian distance (qPCR) and subsequent clustering by UPGMA. Diversity was calculated both per phylum and per the overall microbial composition (by pooling all phyla together). Within-sample diversity was calculated as the Shannon index, that was recently shown to be a robust estimate of microbial diversity. Diversity analysis was performed using the vegan software package in R.

Partial least squares discriminant analysis (PLS-DA)

A partial least squares discriminant analysis (PLS-DA) regression model was used for the prediction of clinical status of samples; i.e. whether it belonged to a parodontitis patient or to a control subject. PLS-DA is a supervised classification method, which aims to find linear transformations of a matrix of predictors and categorical responses so as to maximize their covariance. The PLS-DA model was constructed on the basis of four different datasets: one for each of the three separate phylum groups and one for the overall microbial composition, by pooling all phyla. Under the assumption that the more discriminant variables are the ones with a higher variance, a preliminary variable selection was performed by filtering out low variance predictors. Only the top 25% most variable predictors were considered in the analysis.

The Variable Importance for Projection (VIP) criterion was used to rank the different OTUs based on their contribution to the response variable (clinical status) and PLS components. PLS-DA provides a quantitative estimate of the discriminatory power of each descriptor by means of VIP (variable importance for the projection) parameters. VIP values rank the descriptors by their ability to discriminate different groups

and is therefore considered an appropriate quantitative statistical parameter. Only the OTUs with the highest contribution (VIP score > 1.2) were considered. The OTUs resulting from this selection were translated to most likely bacterial species by comparison to a database consisting of >1500 bacterial species and their associated IS lengths. PLS-DA analysis was performed using the DiscrMiner package in R (version 2.15.2). All data visualizations were performed with the Spotfire software package (TIBCO, Palo Alto, CA, USA).

10 Results

Quantitative real-time PCR

Of the seven paropathogens in the real-time PCR reactions, two, *F.nucleatum* and *P. micra* were detected in both healthy and diseased subjects. The other five were detected in diseased subjects only. *T. forsythia* was most discriminative, being present in all ten diseased samples and in none of the healthy subjects (Figure 13). When loads were taken into account and patients were clustered based on Euclidian distances of the loads of all seven paropathogens combined, it was clear that diseased samples generally harboured higher loads of the one or more of the seven paropathogens. However, clustering of healthy and diseased samples did not show a clear separation into two distinct groups. Rather, a gradual decline in loads of the paropathogens was seen from the diseased group towards the healthy group with no distinct level of separation (figure 14).

25 *IS-pro*

IS-pro results showed presence of bacteria from all phyletic groups for most samples. When a heat map was constructed and clustered, there was no gradual separation as with the qPCR. In contrast, the healthy and diseased samples clearly fell into two distinct clusters (figure 15).

30 Furthermore, it was seen in the heat map that phyla did not show equal

presence in samples from healthy and diseased subjects. Most notably, *Proteobacteria* showed much higher presence in the parodontitis group than in the healthy group. *Bacteroidetes* too showed a higher presence in the parodontitis group. These effects were quantified by a diversity analysis.

5 Shannon diversity index was calculated for all samples and showed significant differences between healthy and diseased samples (figure 16). Difference in diversity was most outspoken for the phylum *Proteobacteria* with a mean Shannon diversity index of 2,38 for parodontitis samples (IQR 0,56) and 0,69 for healthy samples (IQR 1,09). This difference was highly
10 significant as calculated by a students t-test ($p=0.0022$). A significant difference in diversity was also found for *Bacteroidetes* between healthy and diseased samples ($p=0.0048$). For *Firmicutes* Shannon diversity index was similar for both groups. When all phyla were combined, diversity was still significantly different between the two groups ($p=0.0037$).

15 Clustering of heat maps and diversity analyses are both based on whole-profile analysis. To investigate the contribution of individual OTU's to sample differences based on health status, PLS-DA was employed. With PLS-DA, we were able to classify samples into healthy and diseased groups with 100% accuracy. In figure 17 a low dimensional projection of the PLS-
20 DA analysis is depicted. Interestingly, the OTU's with highest VIP scores belonged to the phylum *Firmicutes*, followed by *Bacteroidetes*. Thus, while *Proteobacteria* showed the most outspoken group effect, this was not attributable to single species. In contrast, presence or absence of particular species was the most important predictor of clinical status for the phylum
25 *Firmicutes*, and to a lesser extent *Bacteroidetes*.

Discussion

Whole microbiome analysis of parodontitis samples shows a disease specific microbial pattern, defined by differential presence of specific
30 species of *Firmicutes* and *Bacteroidetes* and an overall increase in diversity

of *Proteobacteria*. With a predictive modeling approach, we show that these characteristics may be harnessed to generate highly accurate clinical predictions.

In this study we employ total microbiota analysis not only as a
5 research tool, but also as a diagnostic tool and compare it with the current
gold standard of parodontitis diagnostics. We used a parodontitis group as
well as a control group that were well defined. The total microbiota data
were analysed by classical methods such as correlation and diversity
analysis and by predictive modeling. The classical methods are well suited
10 for analysis of whole community structure, while the predictive modeling
can assign importance to individual species.

A drawback of the current study was that the study population
was too small to draw final conclusions. The observed effects on total
microbiota should be confirmed in a larger population. A larger population
15 in combination with cross-validation of the PLS-DA will be necessary for
robust evaluation of applicability of IS-pro as a clinical tool for diagnosing
parodontitis.

qPCR is a well validated method that has shown its value in
clinical practice. However, as it can only measure presence of seven distinct
20 species and it does not give information over total community status or
dynamics. IS-pro does give an overview of the total microbial community
present in the parodontium. Differences in diversity were very outspoken for
the different phyla. For *Proteobacteria*, this effect was greatest, varying
from absence of *Proteobacteria* in a subset of the healthy individuals to a
25 highly diverse population in diseased individuals. A very interesting finding
in this context was that the increase in *Proteobacteria* was not attributable
to particular species, but consisted of different species between individuals.
This may suggest a niche opening up in diseased subjects which may be
exploited by different species of *Proteobacteria*, but not by members of the
30 other phyla. It is interesting to note in this context, that an increase in

Proteobacterial presence has been found in various inflammatory processes in the gut (ref). Furthermore, the presence of several distinct species of the phyla *Firmicutes*, *Fusobacteria* and *Bacteroidetes* in parodontitis has been well established (ref) and has been confirmed in this study.

5 Taken together, our data suggests a process in which distinct species are involved in (initiation of) parodontitis. After this, possibly because of the inflammatory process, a niche might open up which can be exploited by members of the phylum *Proteobacteria*.

 This study contributes to a better understanding of the role of the
10 oral microbiota in parodontitis. Furthermore, it demonstrates how microbiota profiling may be applied to improve clinical diagnostics in oral healthcare. While diagnostic tests aimed at a limited set of known paropathogens have been established, analysis of the total microbiota enables measuring presence and relative quantity of all bacteria present in
15 a sample. By applying a predictive model such as PLS-DA to such datasets, many more parameters can be taken into account for clinical prediction than in defined species assays. Not only can these species be taken into account, but by applying weights to the presence and relative quantity of all of the detected species, clinical prediction may become much more accurate.

20 In conclusion, we demonstrate that even in a small sample set, whole microbiota analysis of oral samples can predict clinical status with high accuracy. This study warrants further evaluation of this approach in a larger sample set with a cross-validated predictive modeling approach.

25 **Example 5. Testing fecal samples of Crohn's Disease patients.**

 Fecal samples obtained by rectal swab as described in Example 3 above were obtained from 7 patients suffering from Crohn's Disease, and analysed using the IS-pro technique as described herein in Examples 2-4 using the primers as described for the phyla *Firmicutes*, *Bacteroidetes* and
30 *Proteobacteria* as described herein above. A total of 7 control subjects was

included as a healthy control population. The results of the analysis are displayed in Figure 18 (profile view) and Figure 19 (band view, or heat map). In this Example, and in all other Examples and Figures displayed herein, the data are displayed as profile view and band view. In the profile
5 views. Blue peaks represent the IS regions amplified from the phylum *Firmicutes*, pink/red peaks represent the IS regions amplified from the phylum *Bacteroidetes*, and yellow peaks represent the IS regions amplified from the phylum *Proteobacteria*. At the right side of the graph it is indicated in bars whether the profiles are from diseased (red) or healthy (green)
10 subjects (except in Figs. 44 and 45, wherein healthy individuals in the Shannon index distribution are indicated in red). Marked differences in the peak profiles between diseased and healthy patients are clearly indicated by boxing of the peak regions.

In the band views or heat maps, each column represents a distinct
15 IS profile obtained from one sample/subject. The color intensity corresponds to the peak-height. Although the different phyla are not recognizable in this display, the phylum association of the various bands was used in the UPGMA clustering analysis, separating diseased (red) from healthy (green). In the instances where marked differences in peak profiles could were not
20 discernable by eye from the profile views, UPGMA clustering could clearly distinguish diseased profiles from healthy profiles.

In the present Example, a clear distinction could be made between the fecal microbiomes of patients suffering from the disease versus the healthy control population. Markedly, the IS profile of the
25 *Proteobacteria* group exhibited much higher diversity in the diseased group, with both a higher number of peaks and higher total peak surface area for amplified ITS fragments having nucleotide lengths between 650 and 1000 nucleotides.

30 **Example 6. Testing fecal samples of Ulcerative colitis patients.**

Fecal samples obtained by rectal swab as described in Example 3 above were obtained from 5 patients suffering from ulcerative colitis, and analysed using the IS-pro technique as described herein in Examples 2-4 using the primers as described for the phyla *Firmicutes*, *Bacteroidetes* and

5 *Proteobacteria* as described herein above. A total of 5 control subjects was included as a healthy control population. The results of the analysis are displayed in Figure 20 (profile view) and Figure 21 (band view).

In the present Example, a clear distinction could be made between the fecal microbiomes of patients suffering from ulcerative colitis

10 versus the healthy control population. Markedly, the IS profile of the phylum *Bacteroidetes* showed clear absence of certain species from that phylum the diseased group, notable an absence of amplified ITS fragments having nucleotide lengths between 390 and 450 nucleotides.

15 **Example 7. Testing fecal samples of Eczema patients.**

Fecal samples obtained by rectal swab as described in Example 3 above were obtained from 3 patients suffering from eczema, and analysed using the IS-pro technique as described herein in Examples 2-4 using the primers as described for the phyla *Firmicutes*, *Bacteroidetes* and

20 *Proteobacteria* as described herein above. A total of 5 control subjects was included as a healthy control population. The results of the analysis are displayed in Figure 22 (profile view) and Figure 23 (band view).

In the present Example, a clear distinction could be made between the fecal microbiomes of patients suffering from eczema versus the

25 healthy control population. Markedly, the IS profile of the *Proteobacteria* group exhibited much higher diversity in the healthy control group. Both a reduced number of peaks and lower total peak surface area for amplified ITS fragments having nucleotide lengths between 650 and 1100 nucleotides was found for the eczema group.

Example 8. Testing fecal samples of Colon carcinoma patients and adenoma control subjects.

Fecal samples obtained by rectal swab as described in Example 3 above were obtained from 10 patients suffering from colon carcinoma, and analysed using the IS-pro technique as described herein in Examples 2-4 using the primers as described for the phyla *Firmicutes*, *Bacteroidetes* and *Proteobacteria* as described herein above. A total of 11 control subjects having pre-cancerous adenoma were included as a control population. The results of the analysis are displayed in Figure 24 (profile view) and Figure 25 (band view).

In the present Example, a clear distinction could be made between the fecal microbiomes of patients suffering from colon carcinoma versus the control population. Markedly, the IS profile of the phylum *Firmicutes* exhibited much higher diversity in the colon carcinoma group. A marked increase in the number of peaks for amplified ITS fragments having nucleotide lengths between 200 and 400 nucleotides was found for the colon carcinoma group.

Example 9. Testing paperpoint samples of patients suffering from an peri-implantitis.

Paperpoint samples obtained as described in Example 4 above were obtained from 7 patients suffering from peri-implantitis, and analysed using the IS-pro technique as described herein in Examples 2-4 using the primers as described for the phyla *Firmicutes*, *Bacteroidetes* and *Proteobacteria* as described herein above. A total of 7 healthy control subjects were included as a control population. The results of the analysis are displayed in Figure 26 (profile view) and Figure 27 (band view).

In the present Example, a clear distinction could be made between the paperpoint samples of patients suffering from peri-implantitis versus the control population. Markedly, the IS profile of the phylum

Bacteroidetes (pink/red peaks) with presence of ITS fragments of 300 nucleotides in length was specific for the diseased group, whereas the phylum *Firmicutes* exhibited a shift in profile from species having ITS fragment lengths of around 150-250 nucleotides in the healthy group,
5 towards species having ITS fragment lengths of around 250-350 nucleotides in the diseased group

Example 10. Testing wound samples of patients suffering from an Abscess.

10 A wound sample of a patient suffering from an abscess was obtained and analysed using the IS-pro techniques as described herein. The sample was also cultivated on standard cultivation media in order to detect the presence of any cultivable bacterium. The results are displayed in Figure 28. No bacterial growth was observed on solid cultivation media.
15 Panel B shows the IS profile of the abscess sample, indicating the presence of a single bacterial species of the phylum *Firmicutes*, *Streptococcus cristatus*.

Example 11. Testing fecal samples of Anorexia nervosa patients.

20 A fecal sample was obtained by rectal swab as described in Example 3 above from a patient suffering from anorexia nervosa, and analysed using the IS-pro technique as described herein in Examples 2-4 using the primers as described for the phyla *Firmicutes*, *Bacteroidetes* and *Proteobacteria* as described herein above. Two control subject fecal samples
25 were included as healthy controls. The results of the analysis are displayed in Figure 29 (profile view) and Figure 30 (band view).

In the present Example, a clear distinction could be made between the fecal microbiome of the anorexia nervosa patient versus the control population. Markedly, the IS profile of the phylum *Bacteroidetes*
30 showed distinct peaks at around 230 nucleotides in the anorexia patient,

wherein the diversity of the phylum *Proteobacteria* had markedly decreased in the anorexia patient relative to the control samples.

Example 12. Testing fecal samples of Asthma patients.

5 Fecal samples of two asthma patients were compared to fecal samples of three control subjects using the IS-pro technique as described in the example herein above. The result are depicted in Figure 31 and Figure 32. Asthma patients showed a marked increase in diversity of the *Proteobacteria* group.

10

Example 13. Testing vaginal swab samples of bacterial vaginosis patients.

 Vaginal swab samples were obtained from 6 patients suffering from bacterial vaginosis and were compared with samples obtained from 9
15 control subjects using the IS-pro technique as described in the Examples above. The result are depicted in Figures 33 and 34. A clear distinction could be made between the microbiomes of both populations, whereas the marked difference was in the absence in bacterial vaginosis patients of clear bands in the *Firmicutes* profile at 250, 260 and 520 nucleotides.

20

Example 14. Testing fecal samples of Obstipation patients.

 Fecal samples of patients suffering from obstipation patients were compared with samples of control subjects as described above. The results are depicted in Figures 35-37. A clear distinction could be made between the
25 fecal microbiomes of the two populations. The cumulative profile indicated the presence of one or more *Proteobacterial* species with ITS fragment lengths of around 730 nucleotides was characteristic for the obstipation patients.

Example 15. Testing fecal samples of Irritable Bowel Syndrome patients.

Fecal samples of patients suffering from patients suffering from Irritable Bowel Syndrome were compared with samples of control subjects as described above. The results are depicted in Figures 38-40. A clear distinction could be made between the fecal microbiomes of the two populations. The cumulative profile indicated the presence of several *Bacteroidetes* species with ITS fragment lengths of around 390-450 nucleotides was characteristic for the Irritable Bowel Syndrome patients.

10

Example 16. Testing liquor samples of patients suffering from meningitis.

A cerebro-spinal fluid sample of a patient suffering from meningitis was obtained and analysed using the IS-pro techniques as described herein. The sample was also cultivated on standard cultivation media in order to detect the presence of any cultivable bacterium. The results are displayed in Figure 41. No bacterial growth was observed on solid cultivation media. Panel B shows the IS profile of the liquor sample, indicating the presence of a single bacterial species of the phylum *Firmicutes*, *Streptococcus mitis*.

15
20

Example 17. Testing fecal samples of Rheuma patients.

Fecal samples of patients suffering from patients suffering from rheuma were compared with samples of control subjects as described above. The results are depicted in Figures 42 and 43. Although no marked distinction could be observed in the profile view (figure 42), UPGMA clustering could clearly distinguish diseased profiles from healthy profiles (Figure 43).

25

Example 18. Testing fecal samples of prematurely born children having Necrotising Enterocolitis (NEC).

Necrotizing enterocolitis is an intestinal disorder of prematurely born children, characterized by an acute necrosis of portions of the bowel.

5 The disease has a high mortality rate. Fecal samples of patients suffering from Necrotising Enterocolitis were compared with samples of control subjects as described above. The results are depicted in Figure 44. With a method according to the invention, we can predict at least up to 3 days in advance which children will develop this disease, enabling doctors to
10 potentially prevent the occurrence of the disease before its full onset.

Diagnosis is made by assessing the absence or presence of specific peaks or by applying machine learning algorithms on whole profiles such as Support Vector Machines, Random Forest or partial least squares discriminant analysis (ROC curve; Figure 44C). It was found (Figure 44B)
15 that in samples of Necrotising Enterocolitis (NEC) patients, the abundance and/or species diversity in the phylum of *Firmicutes* decreased as compared to the healthy control population, while the abundance and/or species diversity (as indicated by the Shannon Diversity Index) in the phylum of *Proteobacteria* had increased as compared to the healthy control population.

20 By making an early diagnosis based on microbiota content as described herein, a directed treatment can be applied, either by modulating the microbiota towards a healthy state or by suppressing the host response.

Example 19. Testing fecal samples of neonates having sepsis.

25 Newborn babies, especially those that are prematurely born and/or are admitted to a hospital have a high risk of developing sepsis. Fecal samples of patients suffering from sepsis were compared with samples of control subjects as described above. The results are depicted in Figure 45. With a method according to the invention, sepsis can be predicted at least
30 three days in advance by monitoring of the intestinal microbiota. It is seen

that changes between healthy and diseased patients occur in specific peaks (e.g. those belonging to *Staphylococcus epidermidis*) or at the phylum level (lower abundance or diversity of *Proteobacteria*, *Bacteroidetes*, *Firmicutes* or *Actinobacteria*), or at the profile level. At the profile level, samples can be

5 categorized by e.g. machine learning algorithms as mentioned above. As the causative pathogen can be identified in the microbiota profile (here: *Staphylococcus epidermidis*), a directed antibiotic therapy against that particular pathogen can be initiated before it causes sepsis.

10 **Example 20. Testing fecal samples of *Clostridium Difficile* Associated Diarrhea patients.**

Clostridium Difficile Associated Diarrhea is a disease typically afflicting hospitalized patients with previous or ongoing exposure to antibiotics. The disease has a high morbidity, and specific pathogenic

15 strains may also give a high mortality rate (e.g. *C. difficile* ribotype 027). It has been postulated that *Clostridium difficile* can only cause diarrhoea when the rest of the microbiota has a diminished diversity (caused by antibiotics). With a method according to the invention, we can identify *Clostridium difficile* in fecal samples up to the strain level and at the same

20 time measure the diversity of the fecal microbiota (Figure 46). By combining this information, we can come to an accurate diagnosis, providing doctors with information on the presence of *Clostridium difficile*, the strain type and the status of the microbiota, which taken together determines the diagnosis and treatment modality: when microbiota is severely depleted, a

25 choice could be made for microbial supplementation therapy (e.g. fecal transplantation), when microbiota is less depleted, a conventional antibiotics therapy could be given, while no therapy should be given if *Clostridium difficile* is not found. Finally, the effect of therapy can be

30 monitored: whether bacterial suppletion (e.g. FMT) therapy is successful, whether antibiotic therapy is not depleting the microbiota and other effects.

Claims

1. A method of analyzing the composition of a microbiome based on taxonomic variation in the DNA sequence of the microbial 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the microorganisms in said microbiome, wherein the sequences of conserved DNA regions comprised in the 16S and 23S rRNA sequences flanking said ITS region in the genomic DNA of said microorganisms comprise primer binding sites for amplification of said ITS regions, said method comprising the steps of:
 - a) providing a sample of genomic DNA from a microbiome;
 - b) providing a PCR calibrator system, comprising a set of PCR amplification primers at least one of which primers comprises a label, and a set of at least two PCR calibrators, each PCR calibrator consisting of a DNA fragment comprising a spacer region having a DNA sequence of a given length flanked by upstream and downstream adapter DNA sequences that comprise primer binding sites for binding of said PCR amplification primers wherein said set of PCR amplification primers is for PCR amplifying the spacer region DNA sequence of all PCR calibrators in said set of at least two PCR calibrators, wherein the spacer region DNA sequence comprised in each of said PCR calibrators in said set of at least two PCR calibrators is of a different length, and wherein each PCR calibrator in said set of at least two PCR calibrators is present in equal amount or in a known amount relative to other PCR calibrators in said set;
 - c) adding said set of at least two PCR calibrators from said PCR calibrator system to said sample of genomic DNA;
 - d) performing a PCR amplification reaction on said sample of genomic DNA comprising said set of at least two PCR calibrators using said set of PCR amplification primers from said PCR calibrator system as a first set of amplification primers to amplify and provide amplification products of said ITS region(s) comprised in said set of at least two PCR calibrators, and using at least a second set of PCR amplification primers directed to said flanking

conserved DNA regions to thereby co-amplify and provide amplification products of said ITS regions comprised in said sample of genomic DNA, and;

e) providing a standard curve by determining the PCR amplification efficiency of each of said at least two PCR calibrators from said PCR calibrator system in said PCR amplification reaction of step d) and expressing said PCR amplification efficiency as a function of the length of the DNA sequence of the ITS region;

f) determining the length-specific amplification efficiency for ITS regions of different length comprised in said genomic DNA sample and amplified in step d) using the standard curve as provided in step e);

g) determining the abundance of microbial 16S–23S rRNA internal transcribed spacer (ITS) regions of different length in said microbiome using the length-specific amplification efficiencies determined in step f), and

h) analyzing the composition of a population of microorganisms based on the abundances of ITS regions of different length determined in step g).

2. The method according to claim 1, wherein said standard curve is based on at least five PCR calibrators of different length ranging in length from 50 to 1200 bps.

3. The method according to claim 1 or 2, wherein said spacer region DNA sequence is the sequence of at least a part of a microbial 16S–23S rRNA internal transcribed spacer (ITS) region, and wherein said adapter DNA sequences in said set of at least two PCR calibrators replace the DNA sequences of the conserved DNA regions comprised in the 16S and 23S rRNA gene sequences upstream and downstream of said ITS region in the microbial genome from which said ITS region originates; and wherein said adapter DNA sequences have less than 30% sequence identity with the DNA sequences of said conserved DNA regions.

4. The method according to claim 3, wherein said at least a part of a microbial 16S–23S rRNA internal transcribed spacer (ITS) region is a DNA sequence of at least 20 consecutive nucleotides having at least 95% sequence similarity with a DNA sequence of a 16S–23S rRNA ITS region of a microorganism selected from the phyla Firmicutes, Fusobacterium, Deferribacteres, Spirochaetes, Cyanobacteria, Acidobacteria, Nitrospina, Nitrospirae, Caldithrix, Haloanaerobiales, Verrucomicrobia, Chlamydiae, Planctomycetes, Gemmimonas, Fibrobacteres, Chlorobi, Bacteroidetes, Proteobacteria, Thermotogae, Corprothermobacter, Synergites, Thermodesulfobacteria, Desulfurobacterium, Aquificae, Deinococcus-Thermus, Chloroflexi, Tenericutes and Actinobacteria.

5. The method according to claim 4, wherein said DNA sequence of a 16S–23S rRNA ITS region of a microorganism is selected from the phyla Bacteroidetes, Firmicutes, Actinobacteria, Proteobacteria, Fusobacteria and Verrucomicrobia.

6. The method according to any one of claims 1-5, wherein the set of amplification primers of the PCR calibrator system consists of a single forward primer and a single reverse primer for amplifying all PCR calibrators in the system.

7. The method according to any one of claims 1-6, wherein step d) of performing a PCR amplification reaction on said sample of genomic DNA using at least a set of PCR amplification primers directed to said flanking conserved DNA regions comprises the use of a set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phyla *Firmicutes* and *Bacteroidetes*, and a set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phylum *Proteobacteria*.

8. The method according to claim 7, wherein said set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phyla *Firmicutes* and *Bacteroidetes* comprises the use of:

- a) the forward primer 5'-CTGGATCACCTCCTTTCTAWG-3' comprising a first fluorescent label,
- b) the forward primer 5'-CTGGAACACCTCCTTTCTGGA-3' comprising a second fluorescent label;
- c) and three unlabeled reverse primers 5'-AGGCATCCACCGTGCGCCCT-3'; 5'-AGGCATTACACRTGCGCCCT-3'; and 5'-AGGCATCCRCCATGCGCCCT-3'.

9. The method according to claim 7 or 8, wherein said set of PCR amplification primers for amplifying the 16S–23S rRNA internal transcribed spacer (ITS) regions in the genomic DNA of the phylum *Proteobacteria* comprises the use of:

- a) the forward primer 5'-CCGCCCCGTCACACCATGG-3'
- b) at least one of the reverse primers selected from the group consisting of 5'-AATCTCGGTTGATTTCTTTTCCT-3',
5'-AATCTCGGTTGATTTCTTCTCCT-3',
5'-AATCTCTTTTGATTTCTTTTCCTCG-3',
5'-AATCTCATTTGATGTCTTTTCCTCG-3',
5'-AATCTCTTTTGATTTCTTTTCCTTCG-3',
5'-AATCTCTCTTGATTTCTTTTCCTTCG-3',
5'-AATCTCAATTGATTTCTTTTCCTAAGG-3',

wherein at least one of said primers comprises a fluorescent label.

10. The method according to any one of claims 1-9, wherein said microbiome is the microbiome of a respiratory tract, oral cavity, skin, gastrointestinal tract, or urogenital tract or that of urine, saliva, sputum, pus,

wound fluid, or feces, or the microbiome associated with soil, waterbodies or plants.

11. The method according to claim 10, wherein said microbiome of the respiratory tract, oral cavity, skin, gastrointestinal tract, or urogenital tract or that of urine, saliva, sputum, pus, wound fluid, or feces is of a human body.

12. Use of the method of any one of claims 1-11 in diagnosing or monitoring a digestive tract or gastrointestinal disorder in a patient, wherein the disorder is selected from the group consisting of digestive tract or gastrointestinal disorders associated with inflammatory bowel disease (IBD), diverticulitis, irritable bowel syndrome (IBS), coeliac, lactose intolerance, Necrotising Enterocolitis (NEC), Clostridium Difficile Associated Diarrhea and colorectal cancer, or comorbid disorder selected from the group consisting of attention-deficit/hyperactivity disorder (ADHD), obsessive-compulsive disorder (OCD), anxiety, stress, eating disorder, major depressive disorder (MDD), bipolar disorder, depression, and schizophrenia, wherein said use comprises performing said method on a sample of an intestinal flora.

13. The use of claim 12, wherein said sample of the intestinal flora is a sample obtained from a small or large intestine.

14. The use of claim 13, wherein said small or large intestine is a colon or a rectum.

15. The use of claim 13 or 14, wherein said sample of the intestinal flora is in the form of a biopsy specimen of gastrointestinal mucosa or in the form of a feces sample.

16. The use of any one of claims 13-15, wherein said sample of the intestinal flora is obtained using a rectal swab.

17. The use of any one of claims 12-16, wherein IBD is selected from Crohn's disease, colitis ulcerosa, collagenous colitis, lymphocytic colitis, ischemic colitis, microscopic colitis, Behçet's syndrome, diversion colitis, diverticular colitis, eosinophilic colitis, and radiation colitis.

18. Use of the method of any one of claims 1-11 in diagnosing or monitoring a systemic disease in a patient, said use comprising performing said method on a body sample of said patient.

19. The use of claim 18, wherein said systemic disease is selected from the group consisting of arthritis, sarcoidosis, mixed connective tissue disease, spondylitis ankylopoetica, osteoporosis, juvenile idiopathic arthritis, osteoarthritis, rheumatoid arthritis, sepsis, Behçet's disease, Sjögren's syndrome, fibromyalgia, sclerodermia, Raynaud's phenomenon, and systemic lupus erythematosus, and wherein said use comprises performing said method on a sample of an intestinal flora.

20. Use of the method of any one of claims 1-11 in diagnosing or monitoring a skin disorder or chronic wound in a patient, wherein said use comprises performing said method on a sample of the skin or wound fluid.

21. The use of claim 20, wherein said skin disorder is psoriasis, eczema, acne, or rosacea, or wherein said chronic wound is a diabetic ulcer, or ulcer associated with vascular disease.

22. Use of the method of any one of claims 1-11 in diagnosing or monitoring bacterial vaginosis, wherein said use comprises performing said method on a sample of a vaginal flora.

23. Use of the method of any one of claims 1-11 in diagnosing or monitoring an oral, nasal or oropharyngeal disorder selected from

periodontitis, periimplantitis, and oro-nasopharyngeal carcinoma, wherein said use comprises performing said method on a sample of an oral flora.

24. Use of the method of any one of claims 1-11 in diagnosing or monitoring a disorder that affects the upper or lower respiratory tract, wherein the disorder is selected from respiratory infection, chronic obstructive pulmonary disease (COPD), asthma, cystic fibrosis and lung cancer, wherein said use comprises performing said method on a pulmonary sample selected from bronchoalveolar lavage, a sputum sample and a lung biopsy.

25. Use of the method of any one of claims 1-11 in diagnosing or monitoring a microbial infection, wherein said use comprises performing said method on a liquor sample, a pleural sample, a blood sample, a urine sample, an abscess sample, or a tissue sample of an organ.

26. Use of the method of any one of claims 1-11 in determining fertility or the probability of success of an *in vitro* fertilization or embryo transfer procedure in a female, wherein said use comprises performing said method on a sample of a vaginal flora.

27. Use of the method of any one of claims 1-11 in determining the microbial status of a sample, wherein said sample is from a source selected from environmental, plant, animal and food origin or a sample of a pharmaceutical or chemical product intended to be devoid of microbes or microbial DNA, wherein said use comprises performing said method on a sample from said source, and wherein said status is indicative of sterility of said sample, quality of the environment, microbial safety of food, or the health of plant or animal.

1/55

Fig. 1A

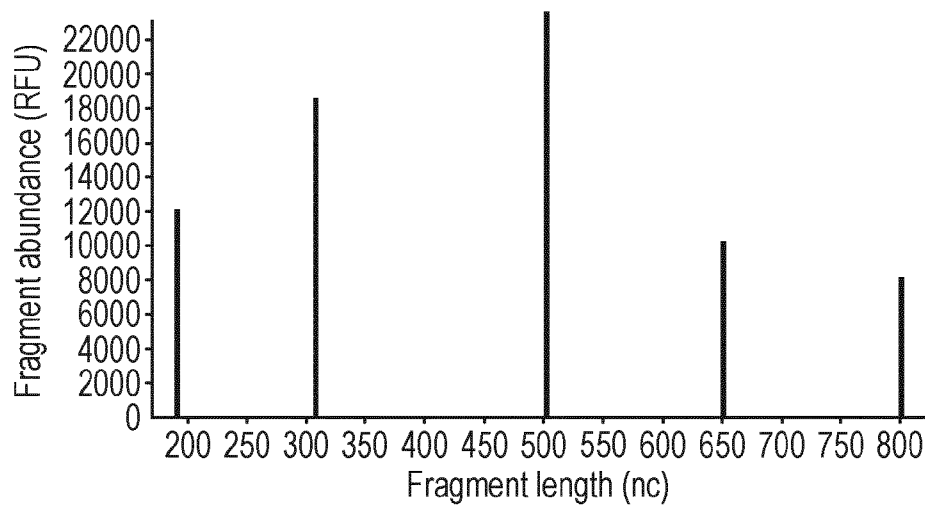


Fig. 1B

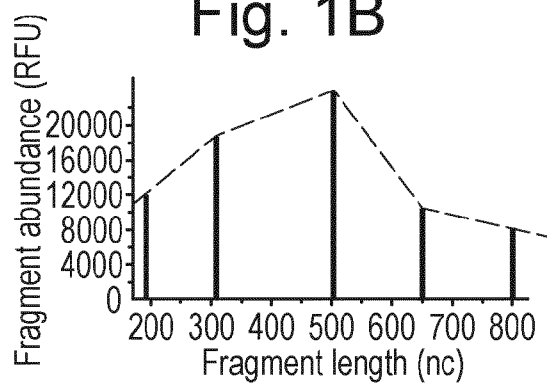


Fig. 1C

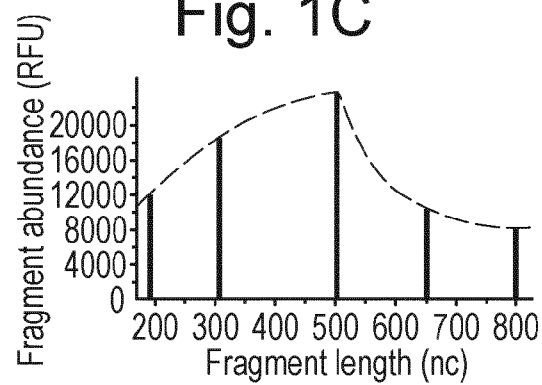


Fig. 1D

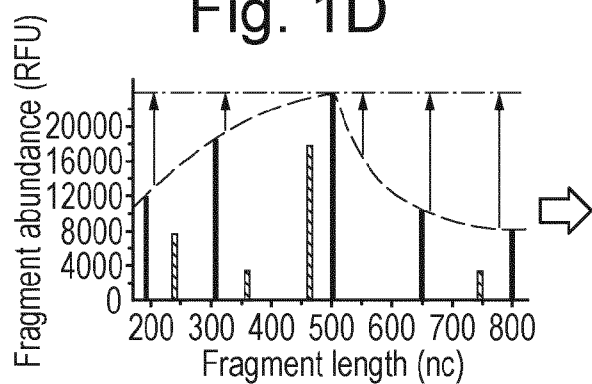
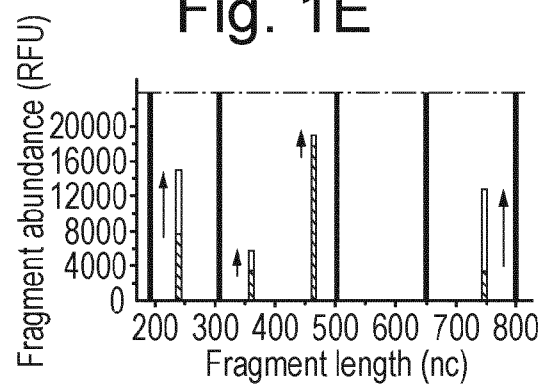
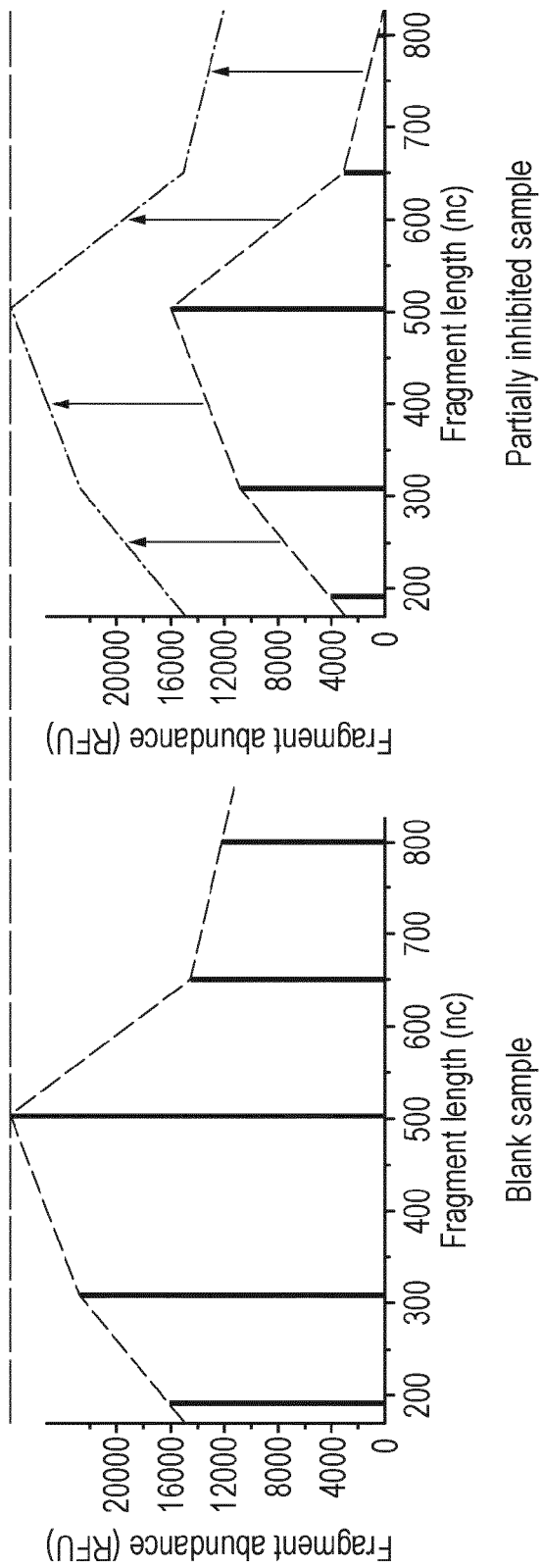


Fig. 1E



2/55

Fig. 2



3/55

Fig. 3A

red peak pattern

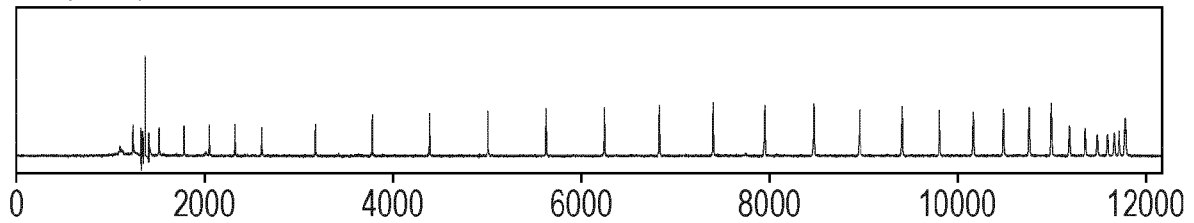


Fig. 3B

red peak pattern + length assignment

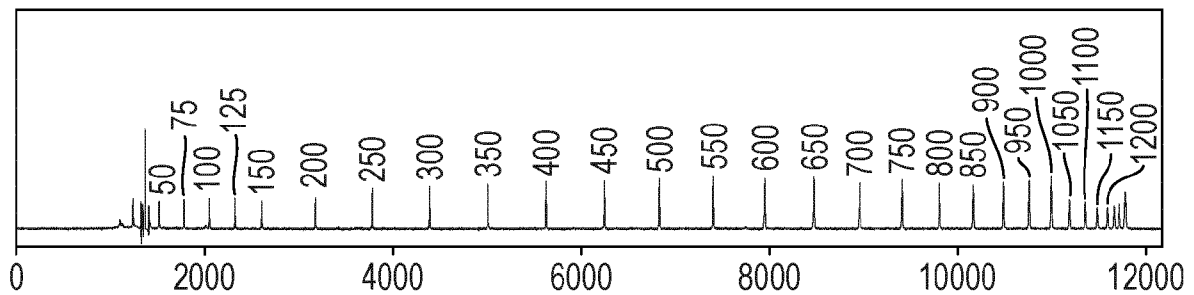


Fig. 3C

red peak pattern + length assignment, + calibrator length

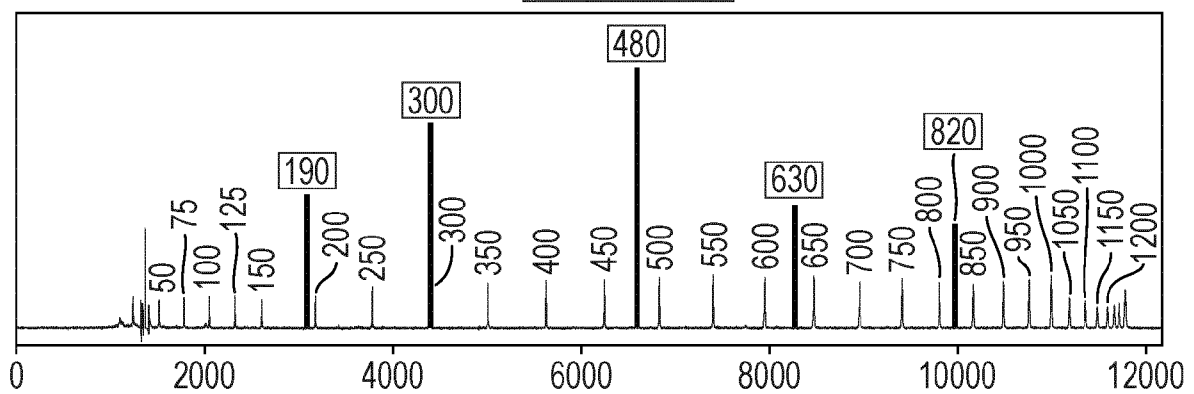
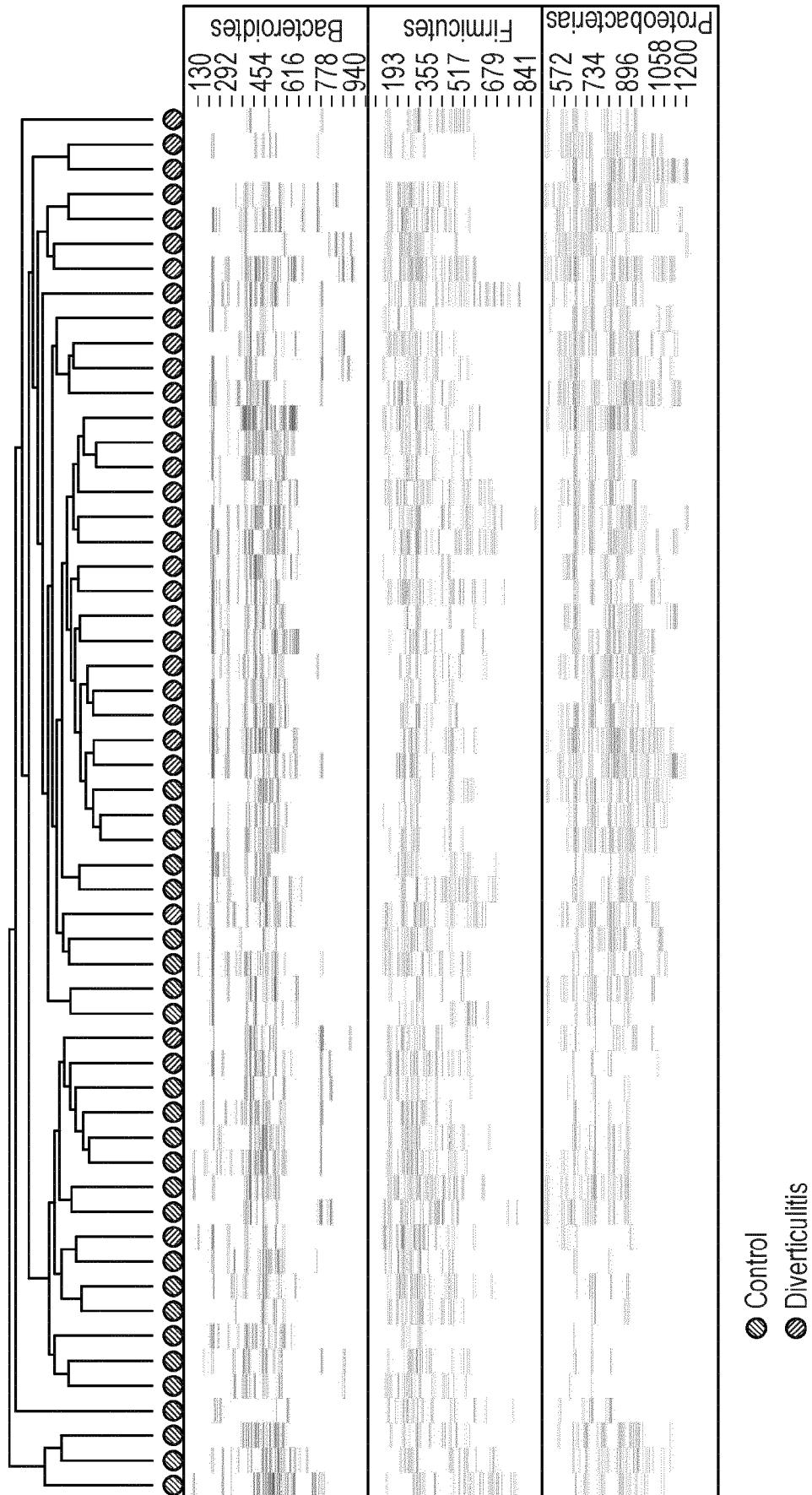
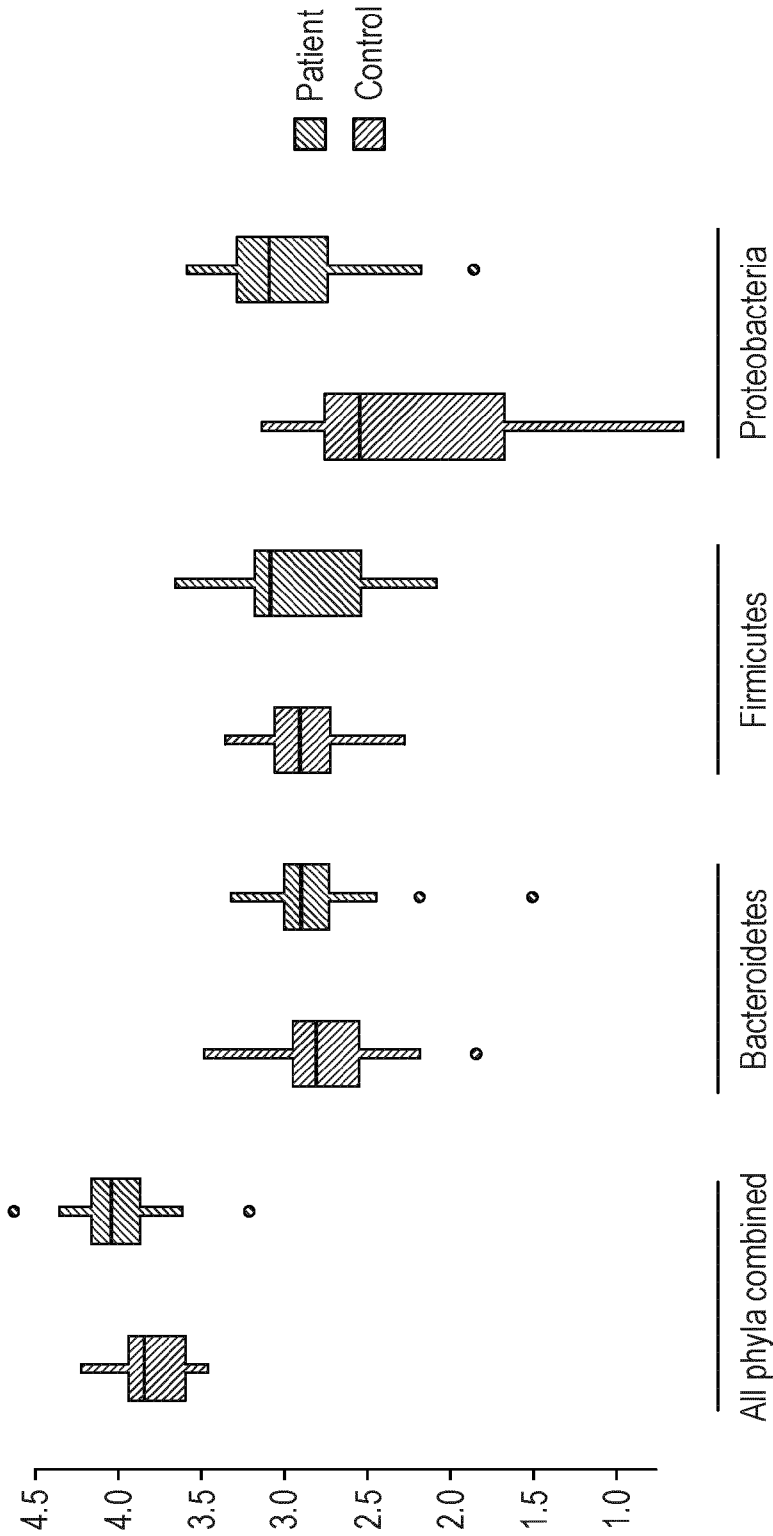


Fig. 4



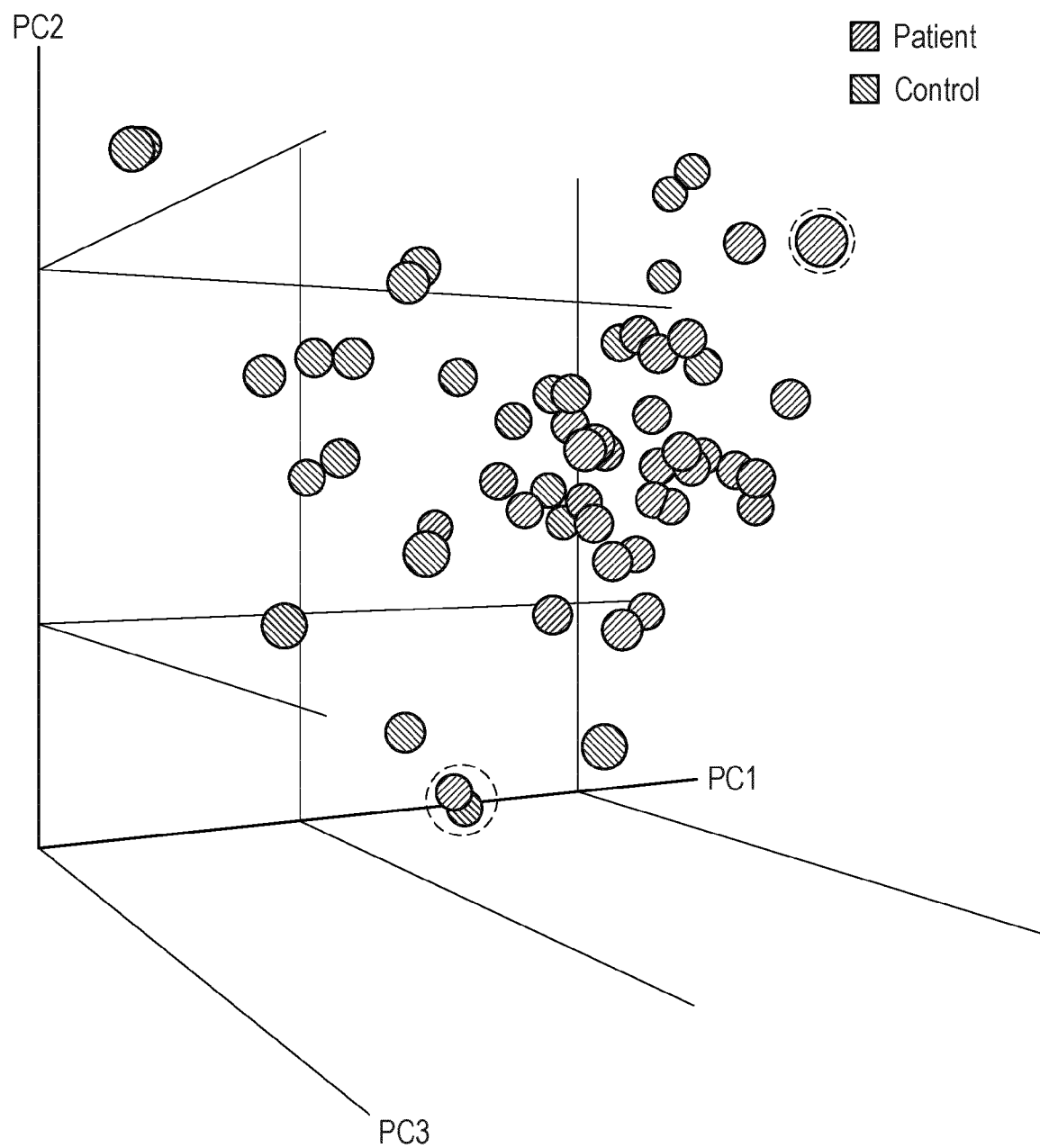
5/55

Fig. 5



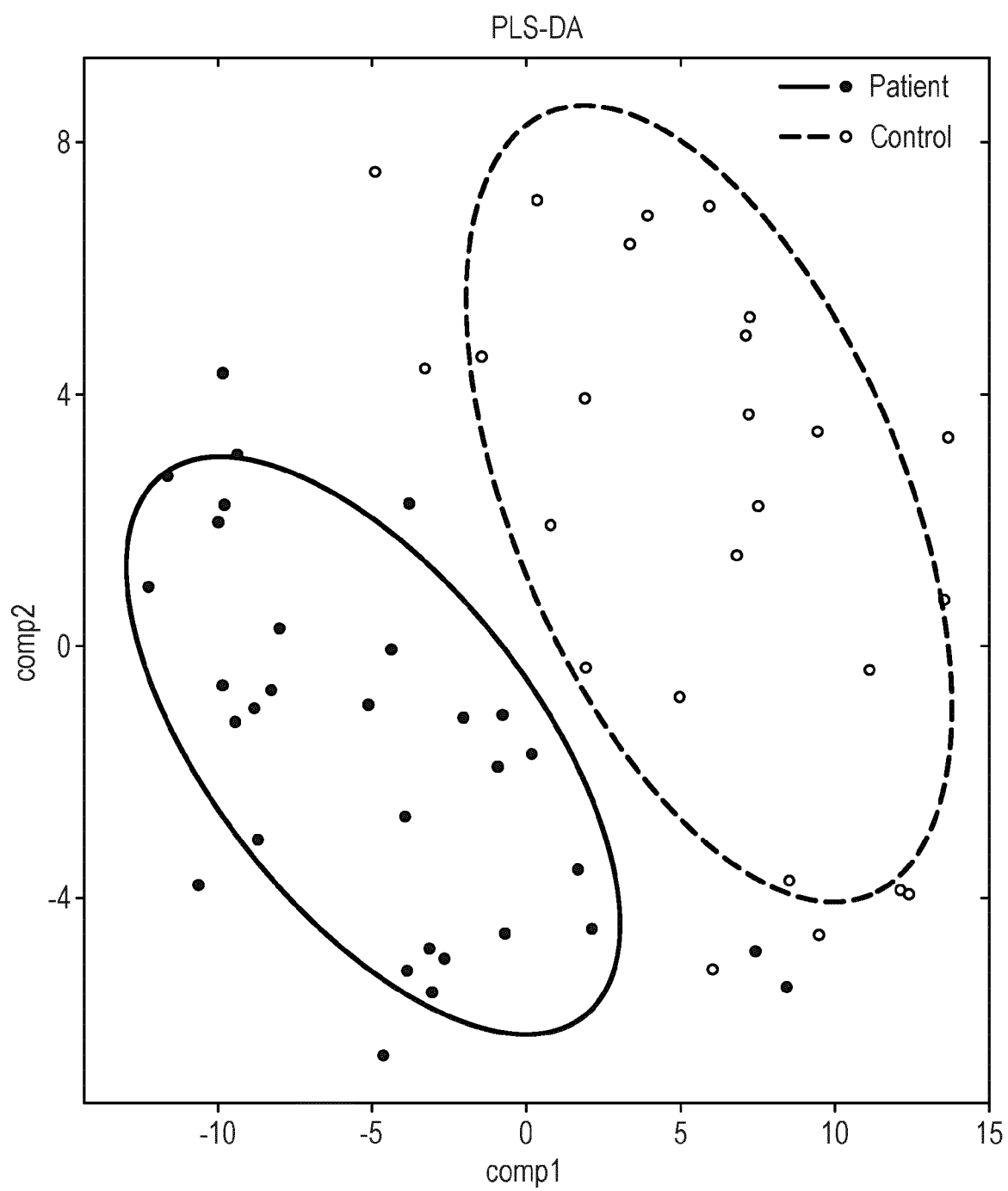
6/55

Fig. 6



7/55

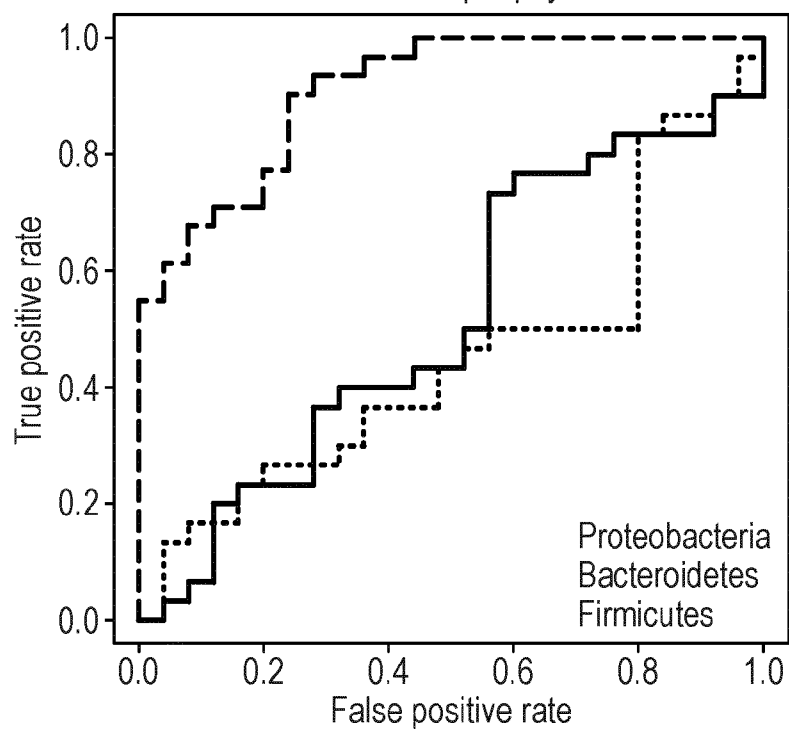
Fig. 7



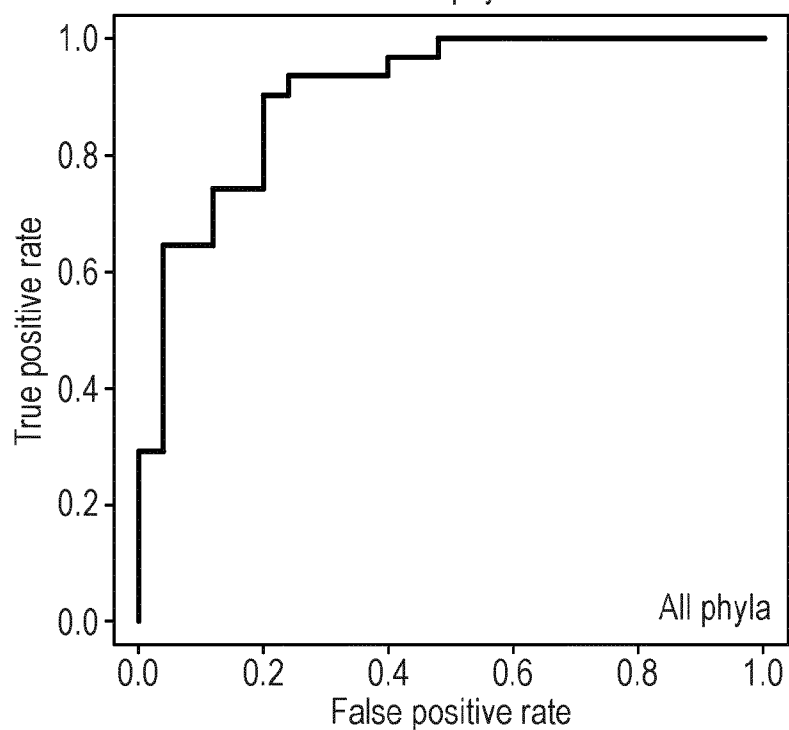
8/55

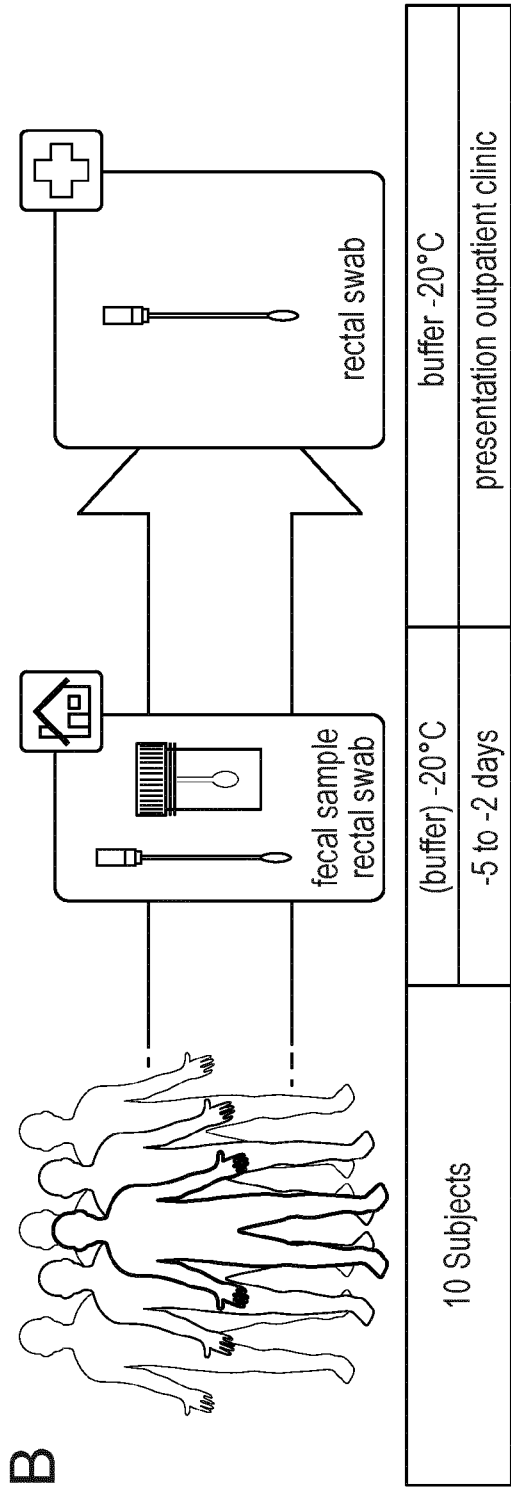
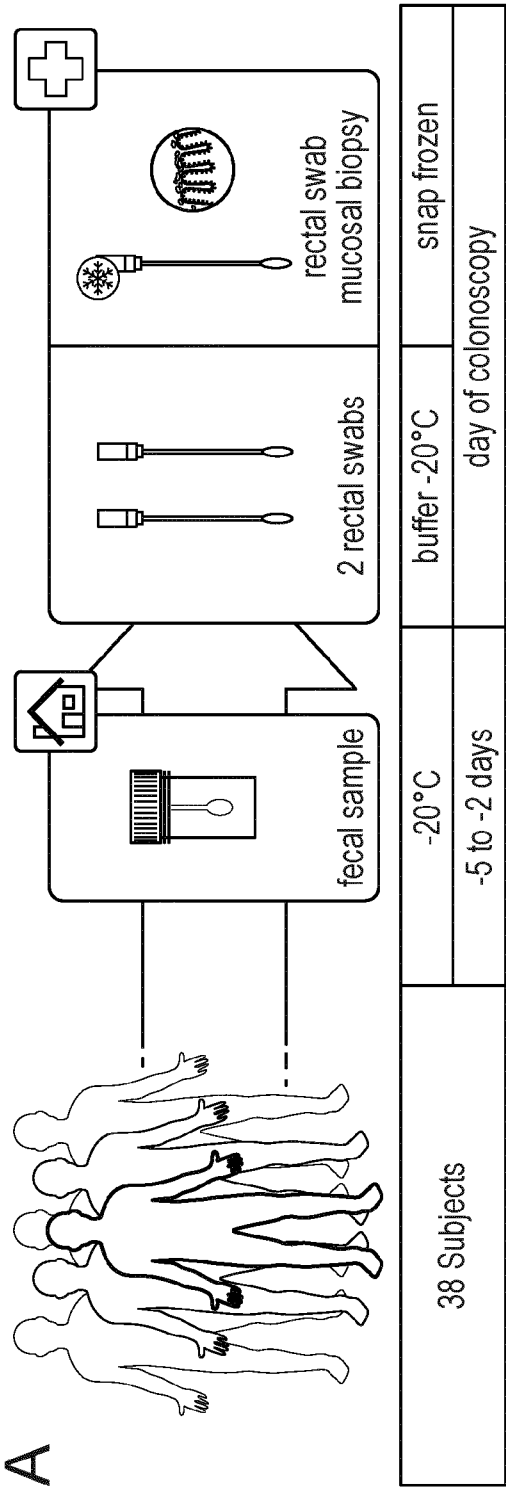
Fig. 8A

ROC curve per phylum

**Fig. 8B**

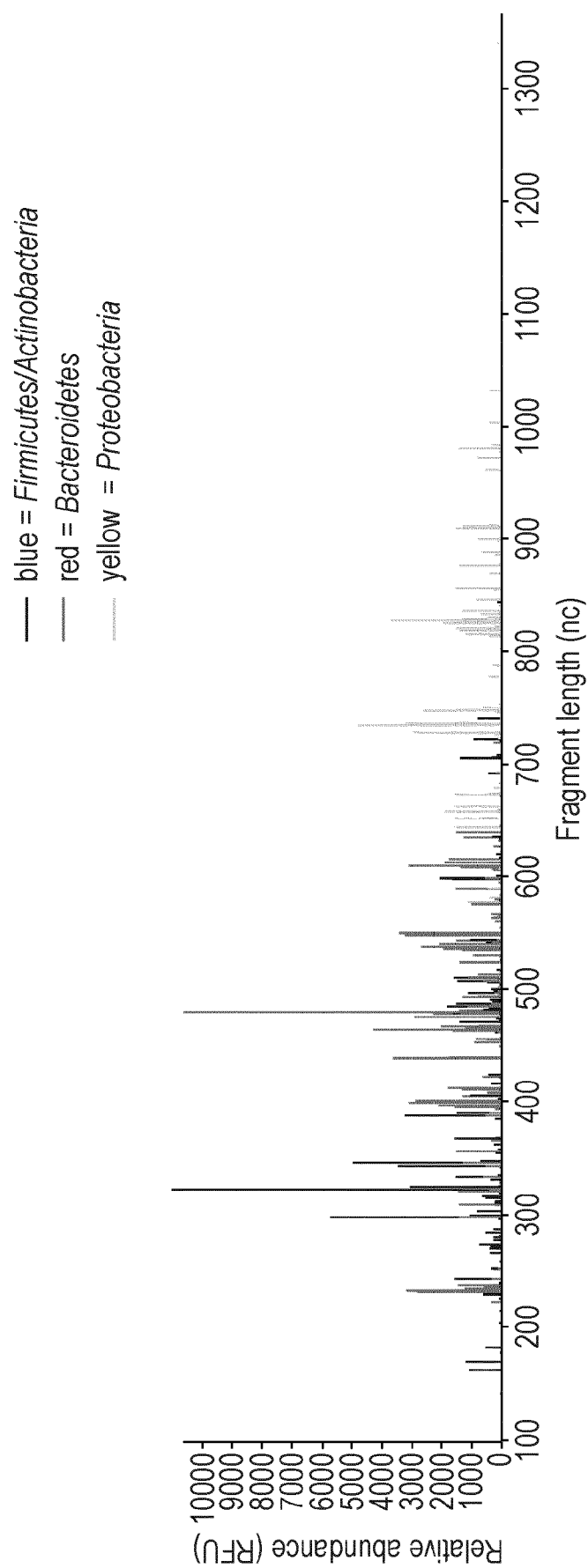
ROC curve all phyla combined





10/55

Fig. 10



11/55

Fig. 11A

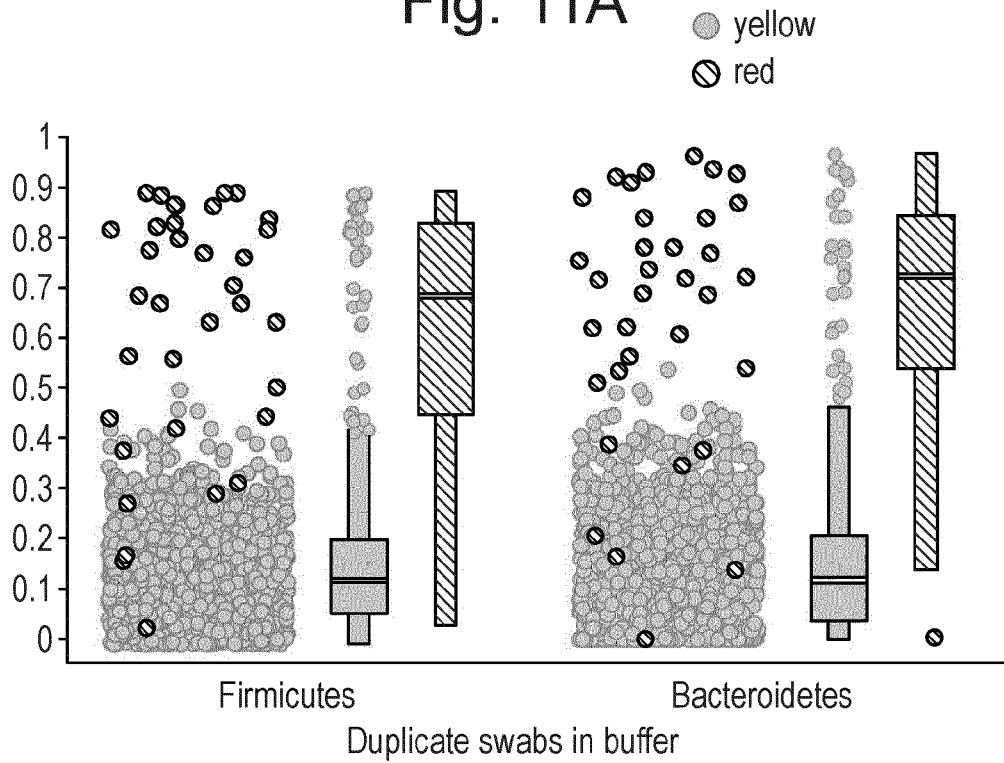
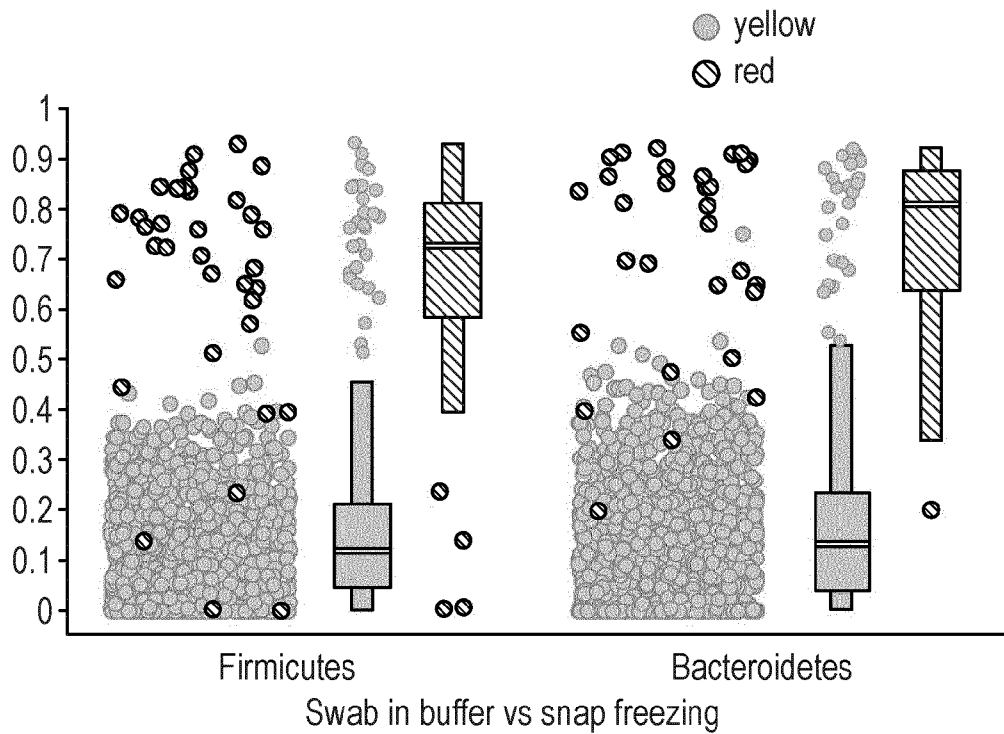


Fig. 11B



12/55

Fig. 11C

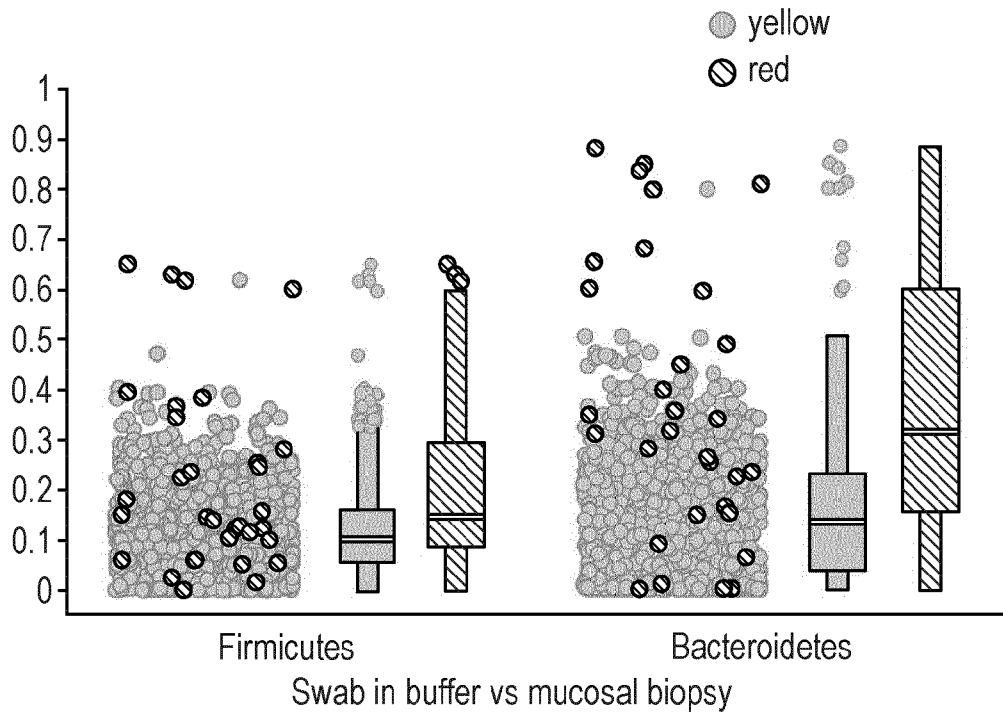
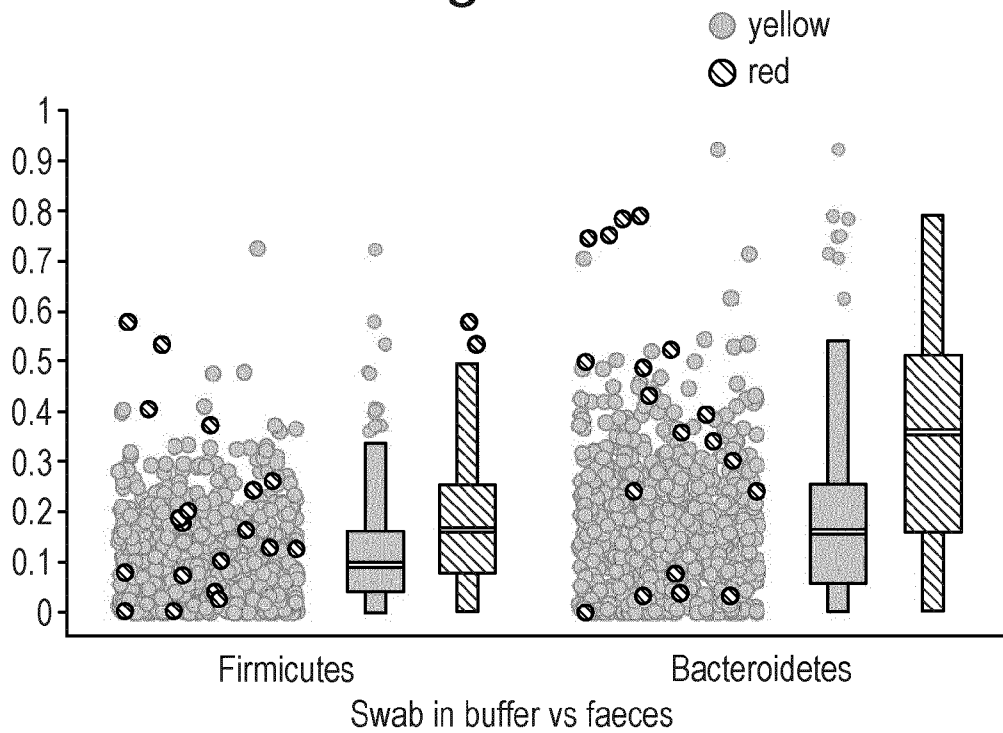
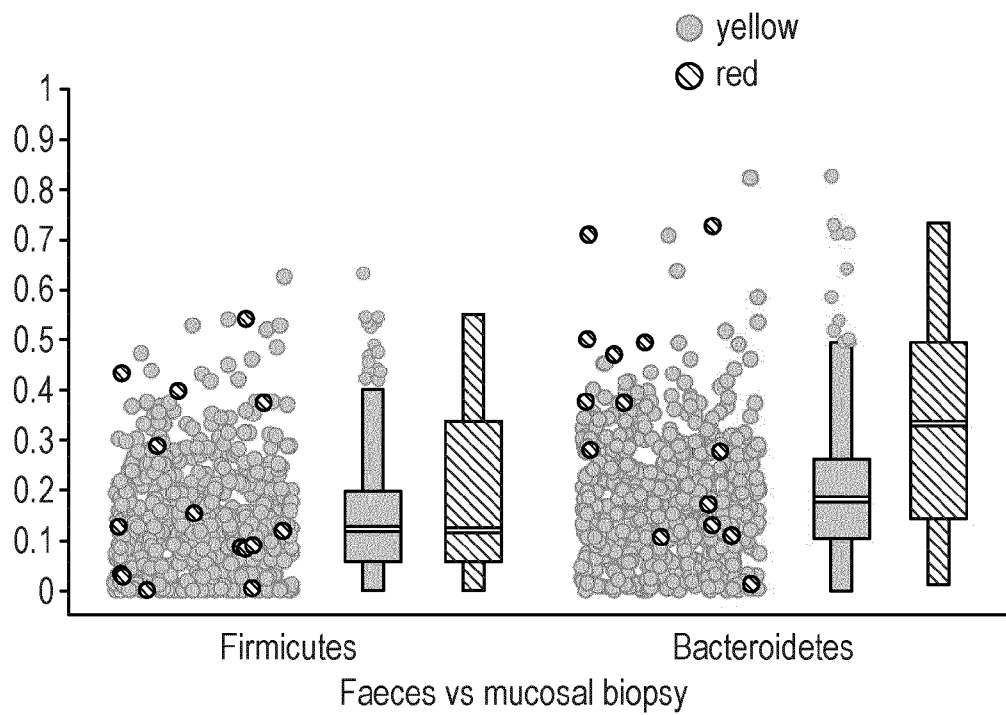


Fig. 11D



13/55

Fig. 11E



14/55

Fig. 12A

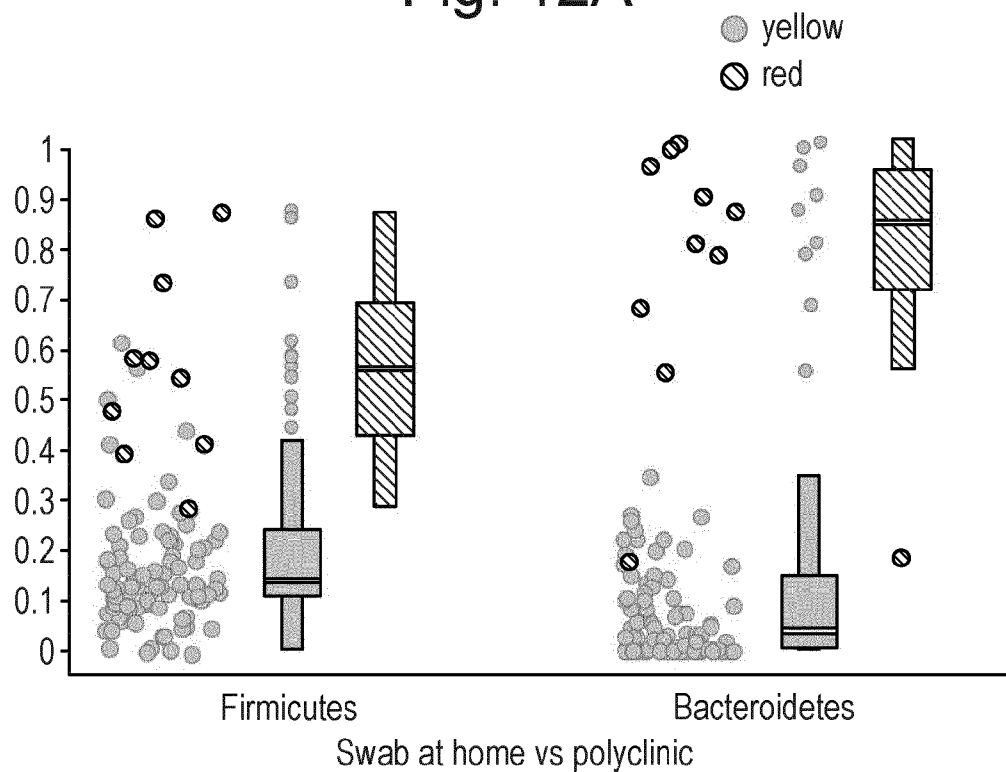


Fig. 12B

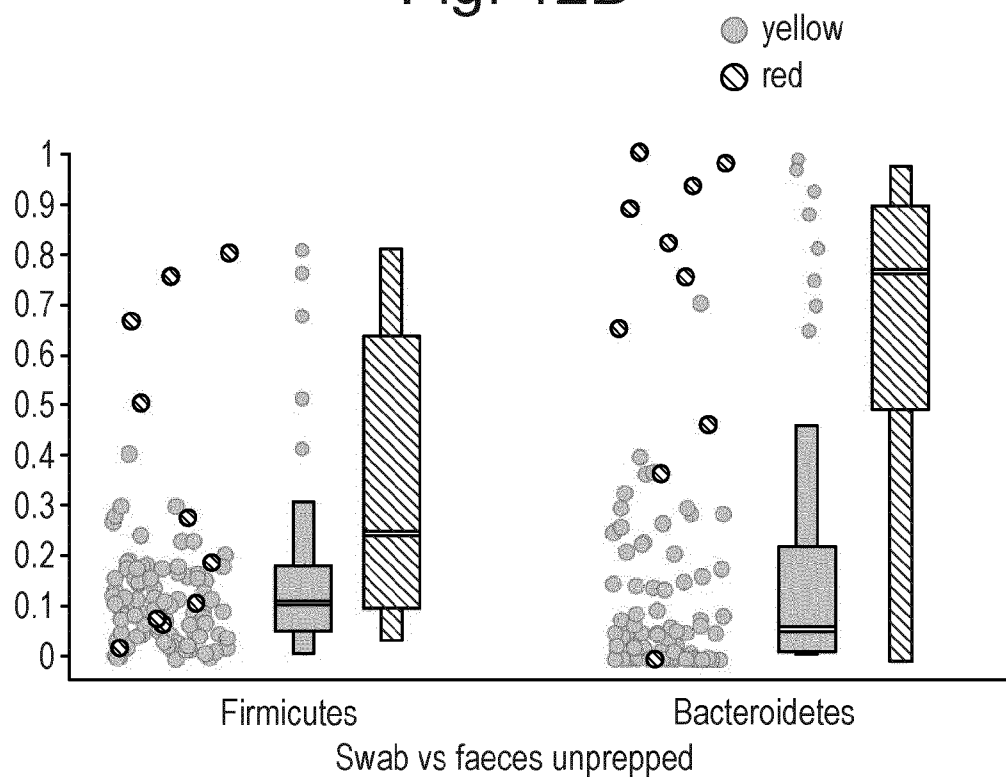
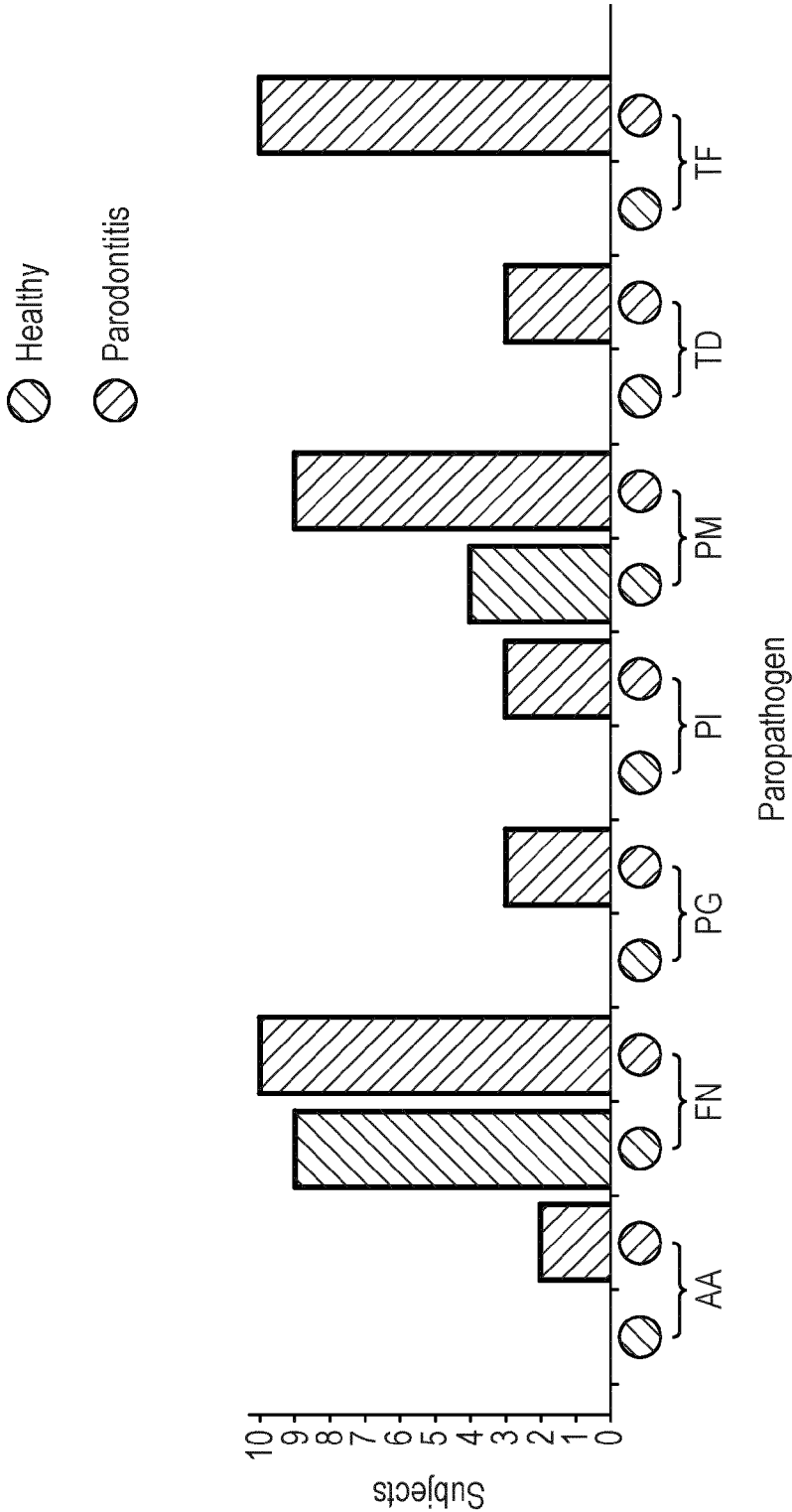
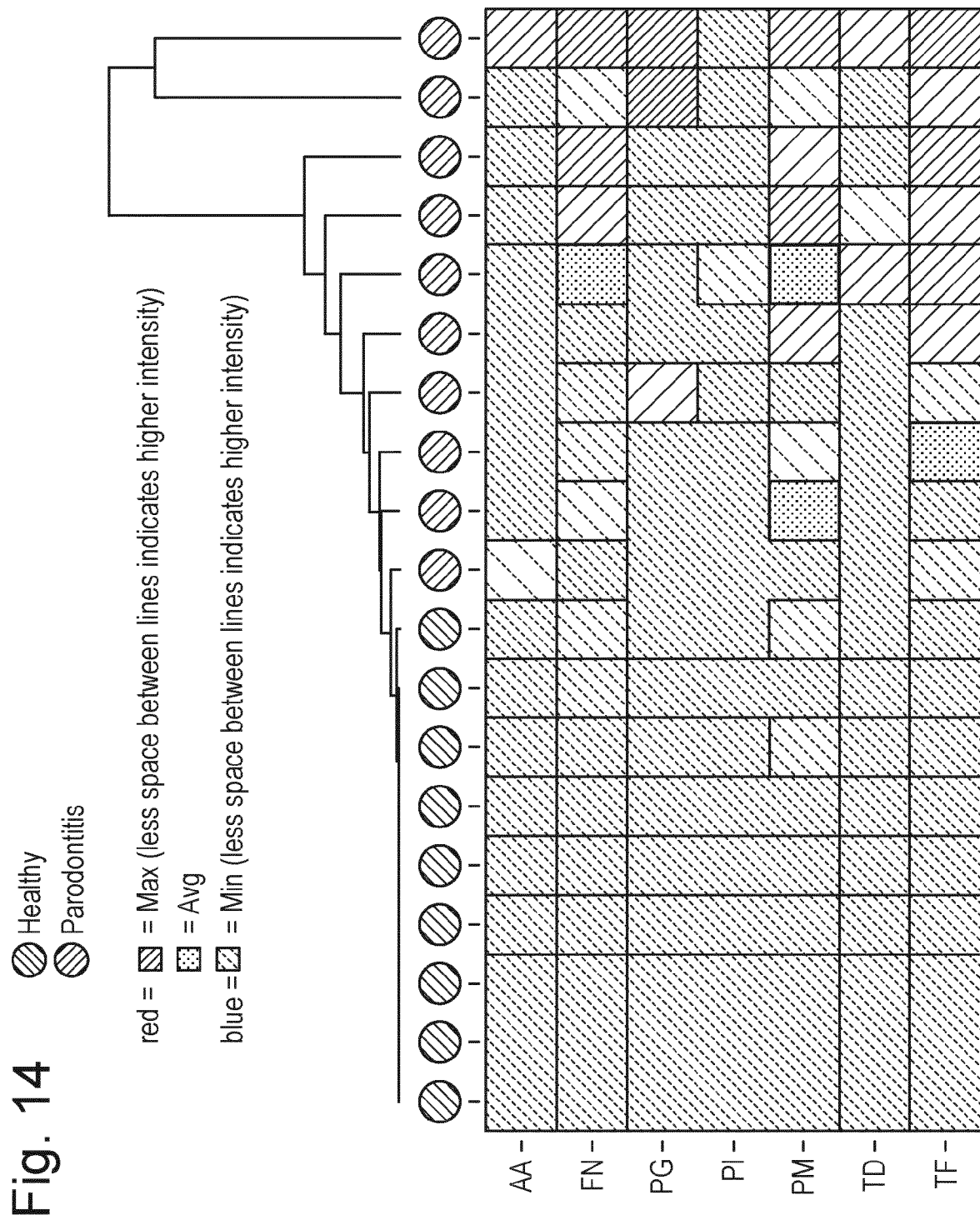
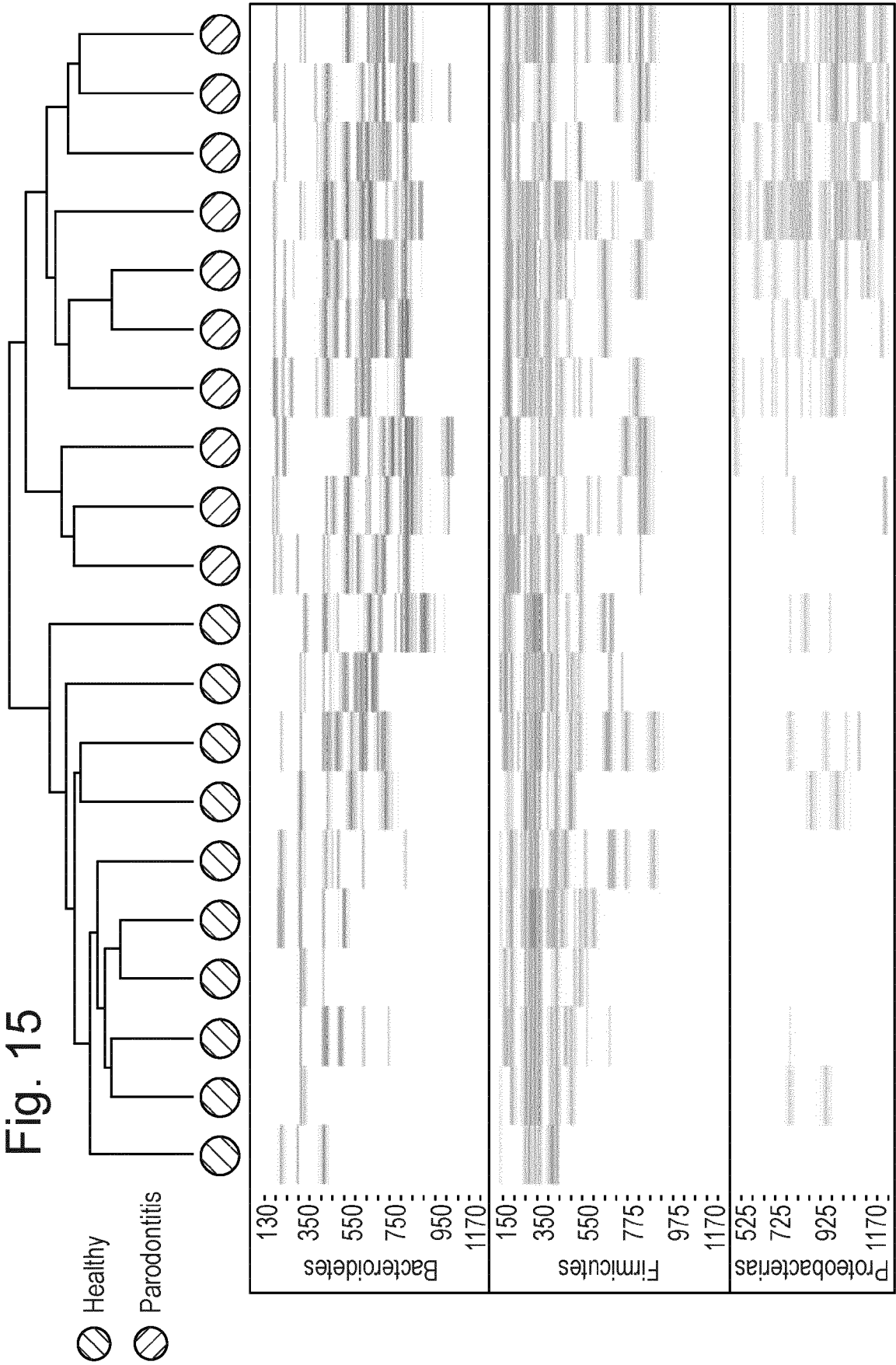


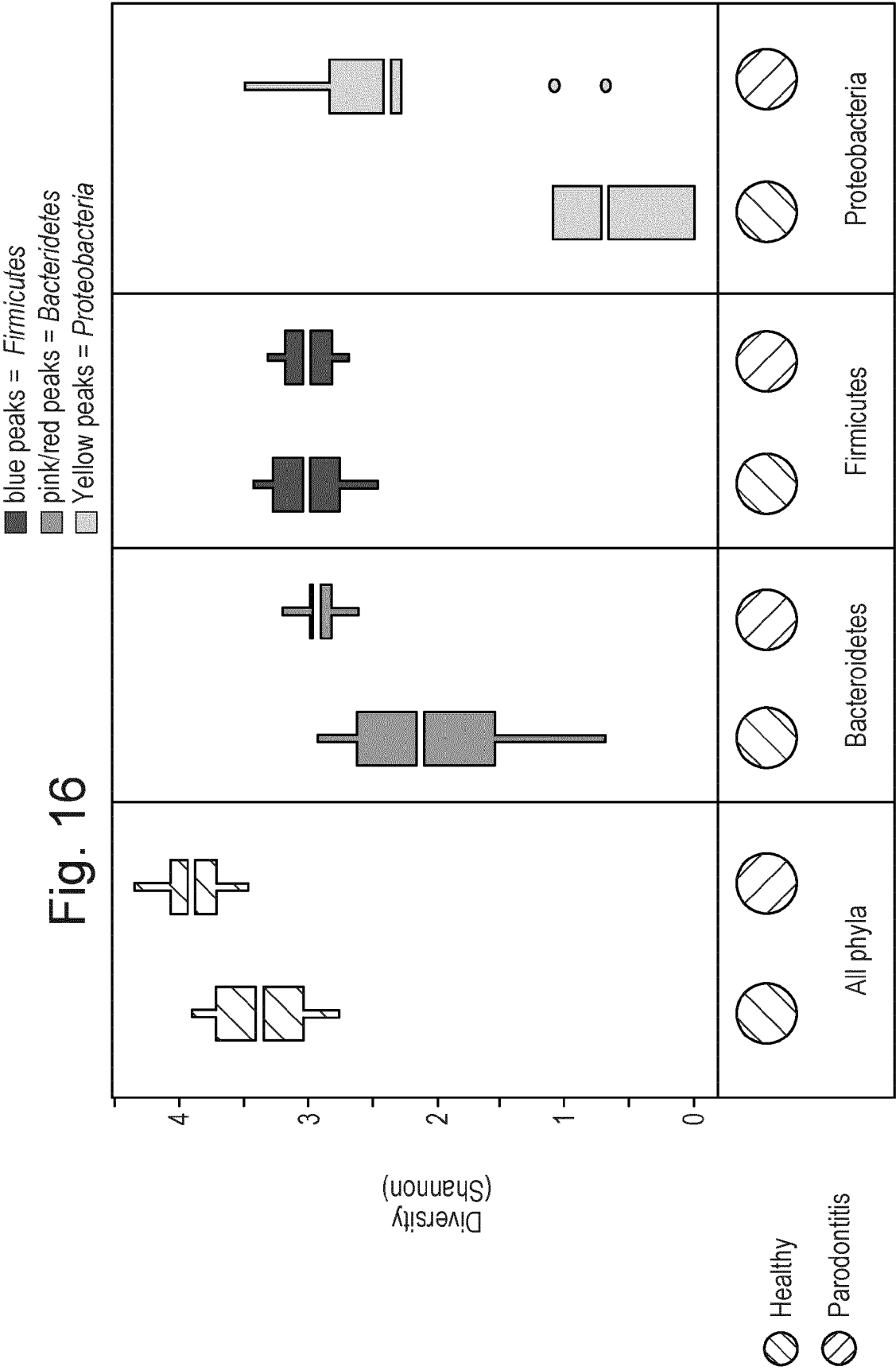
Fig. 13



16/55

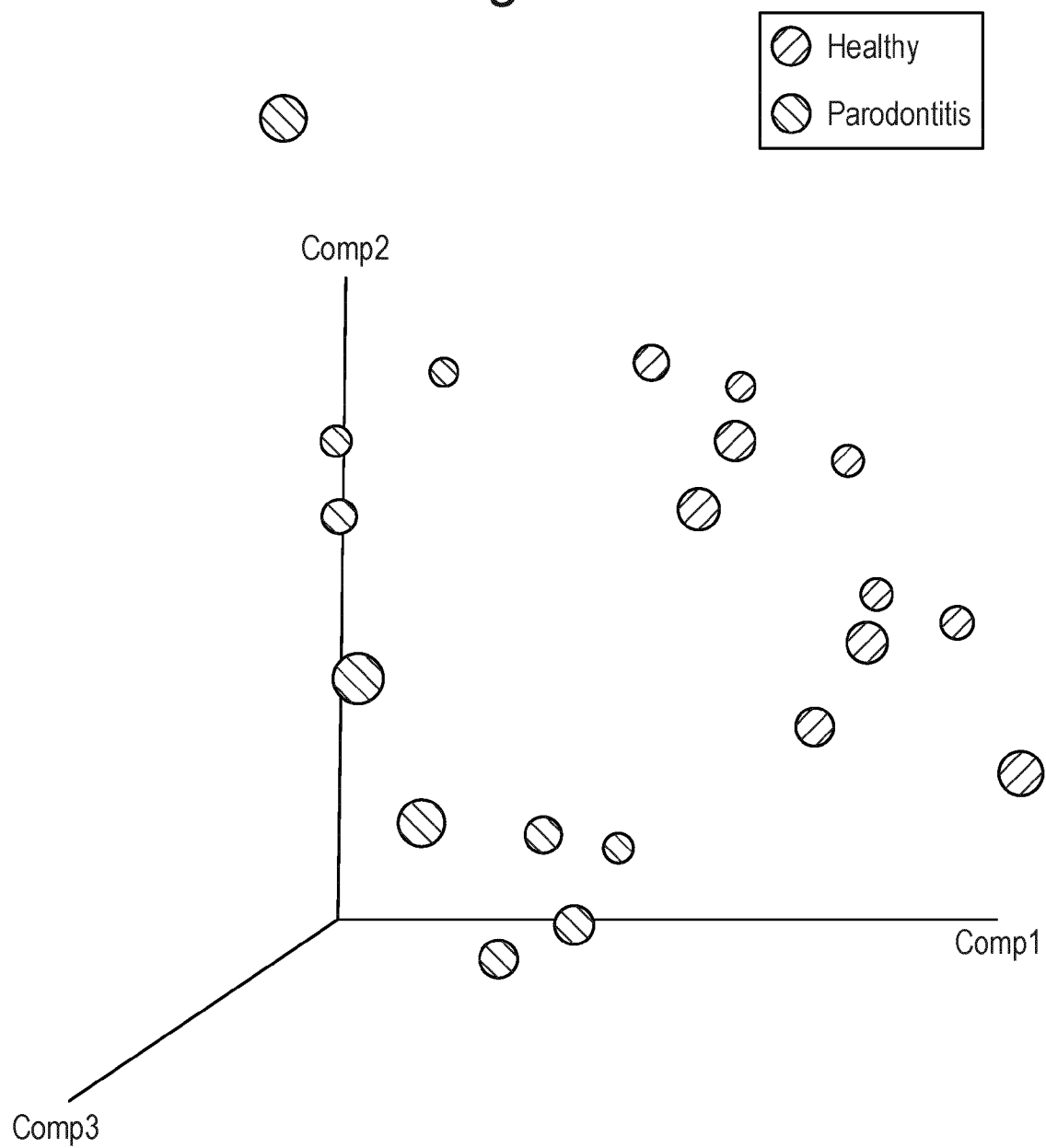




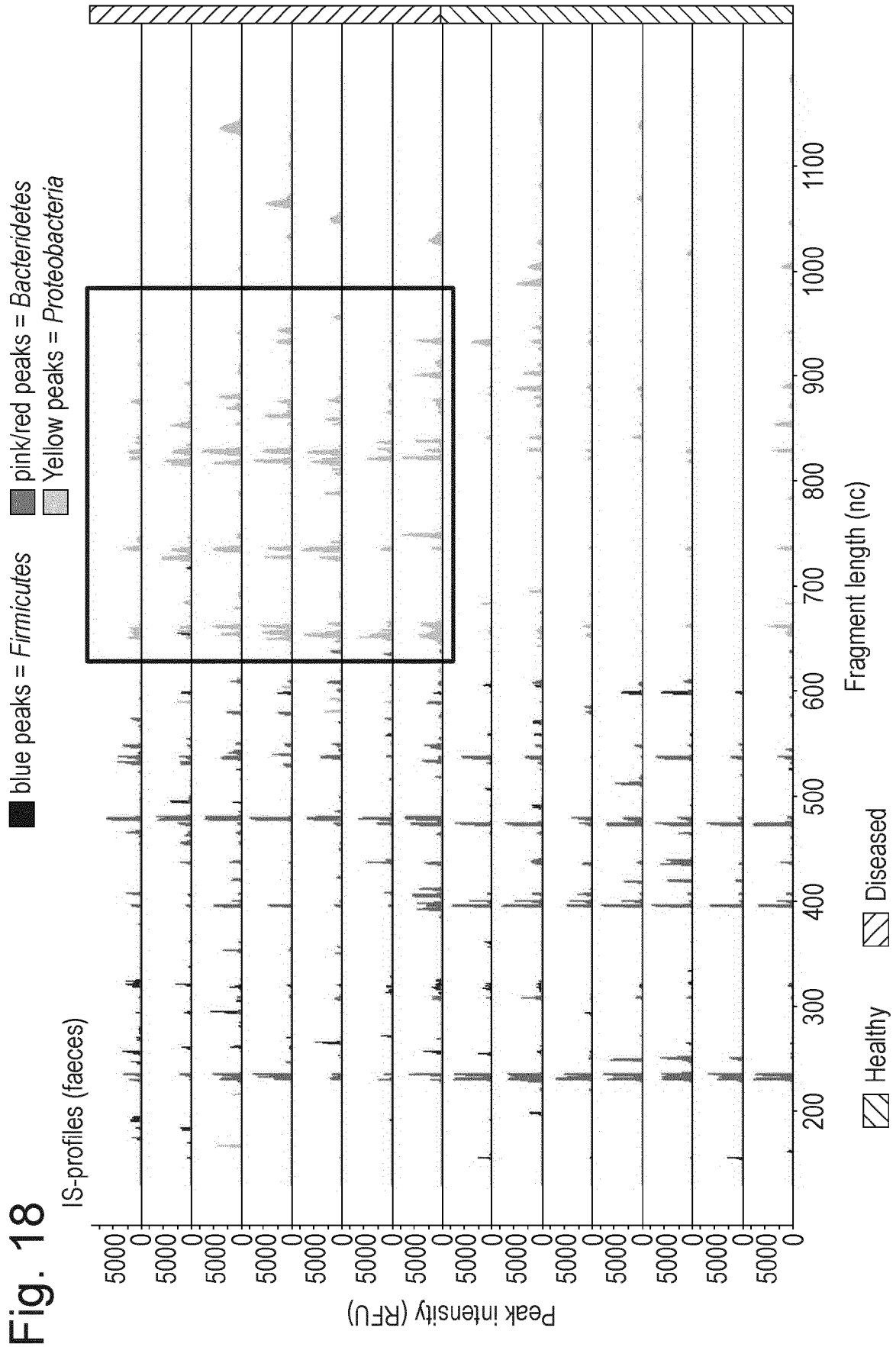


19/55

Fig. 17



20/55



21/55

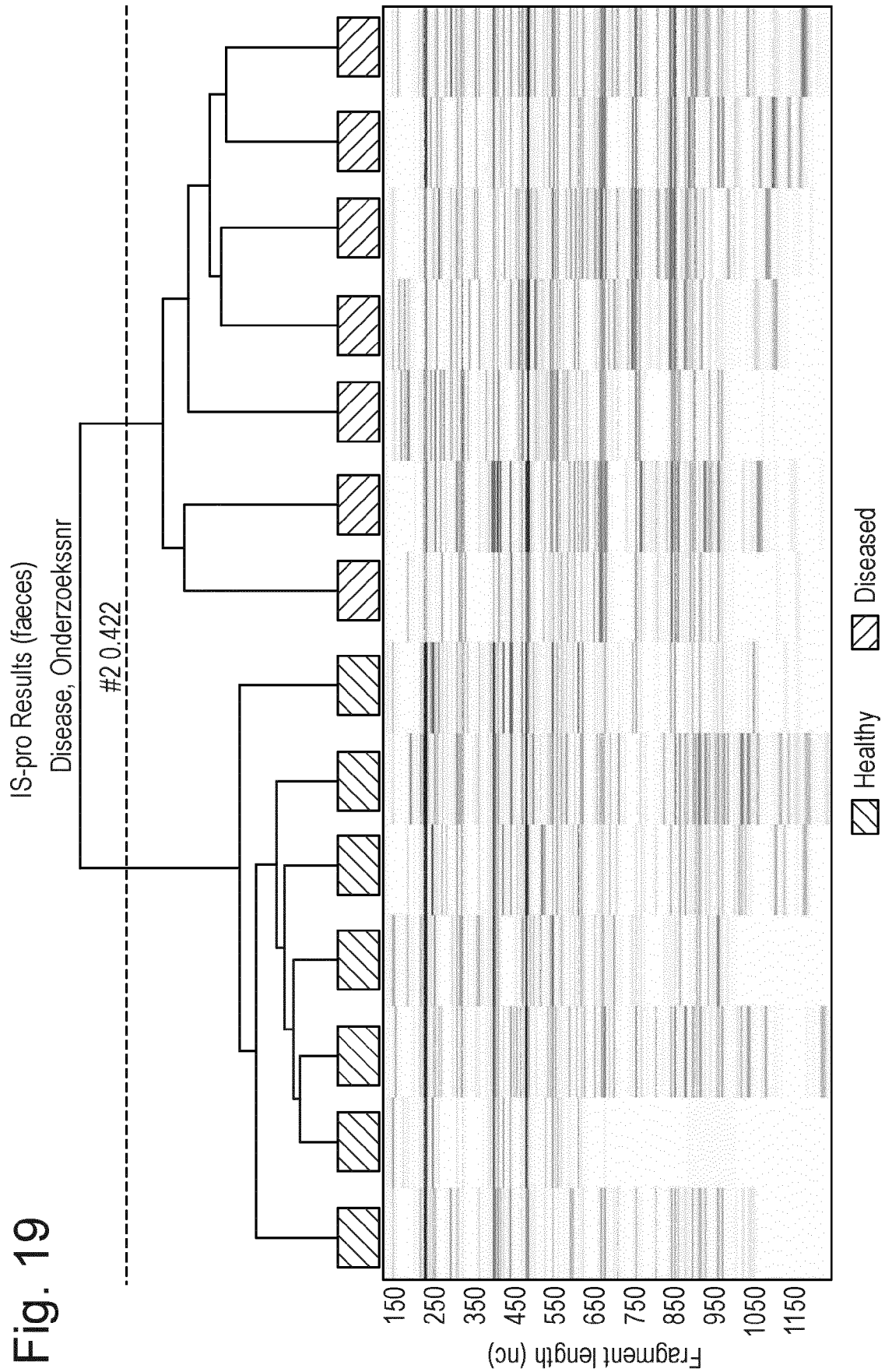
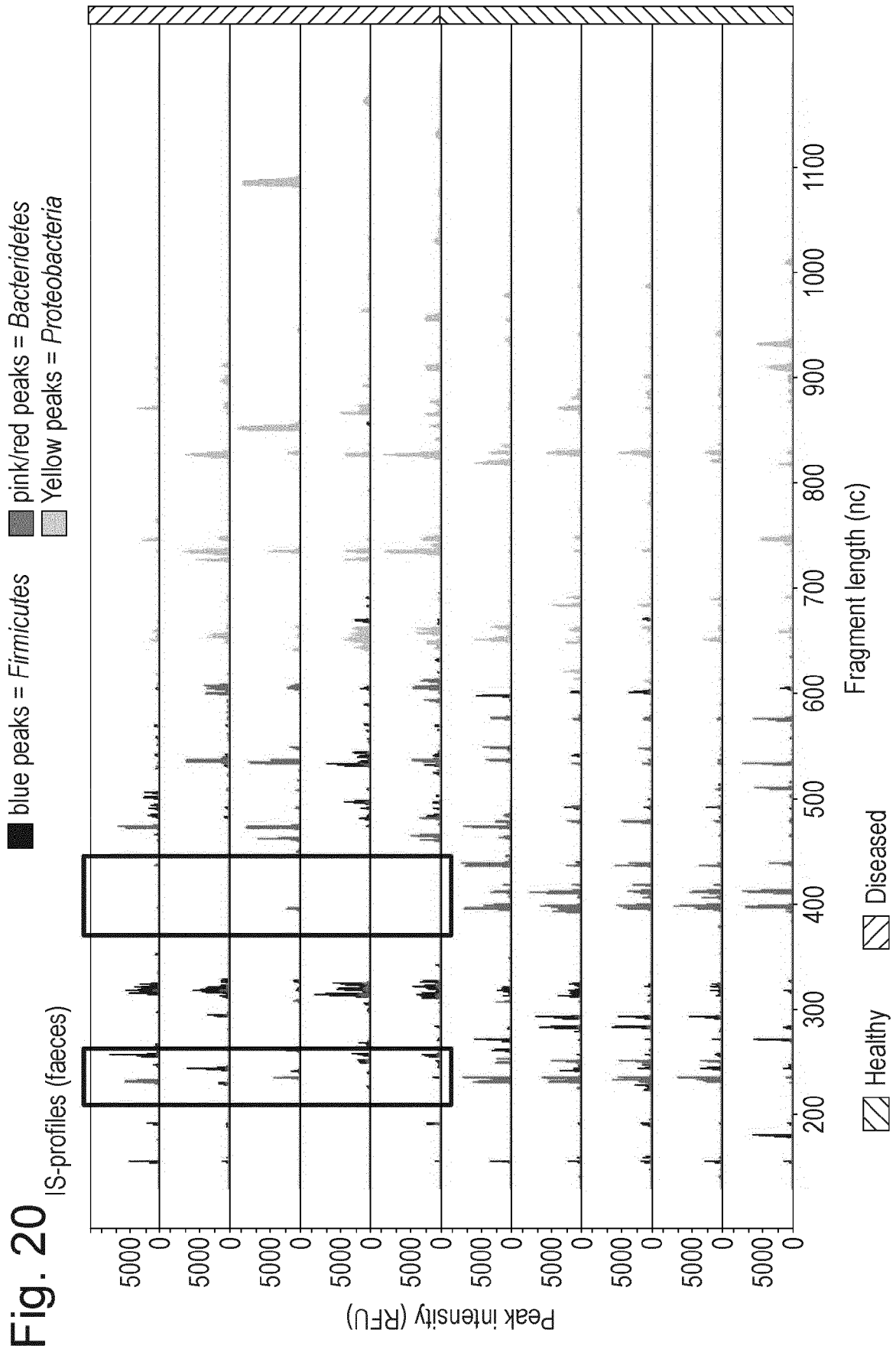
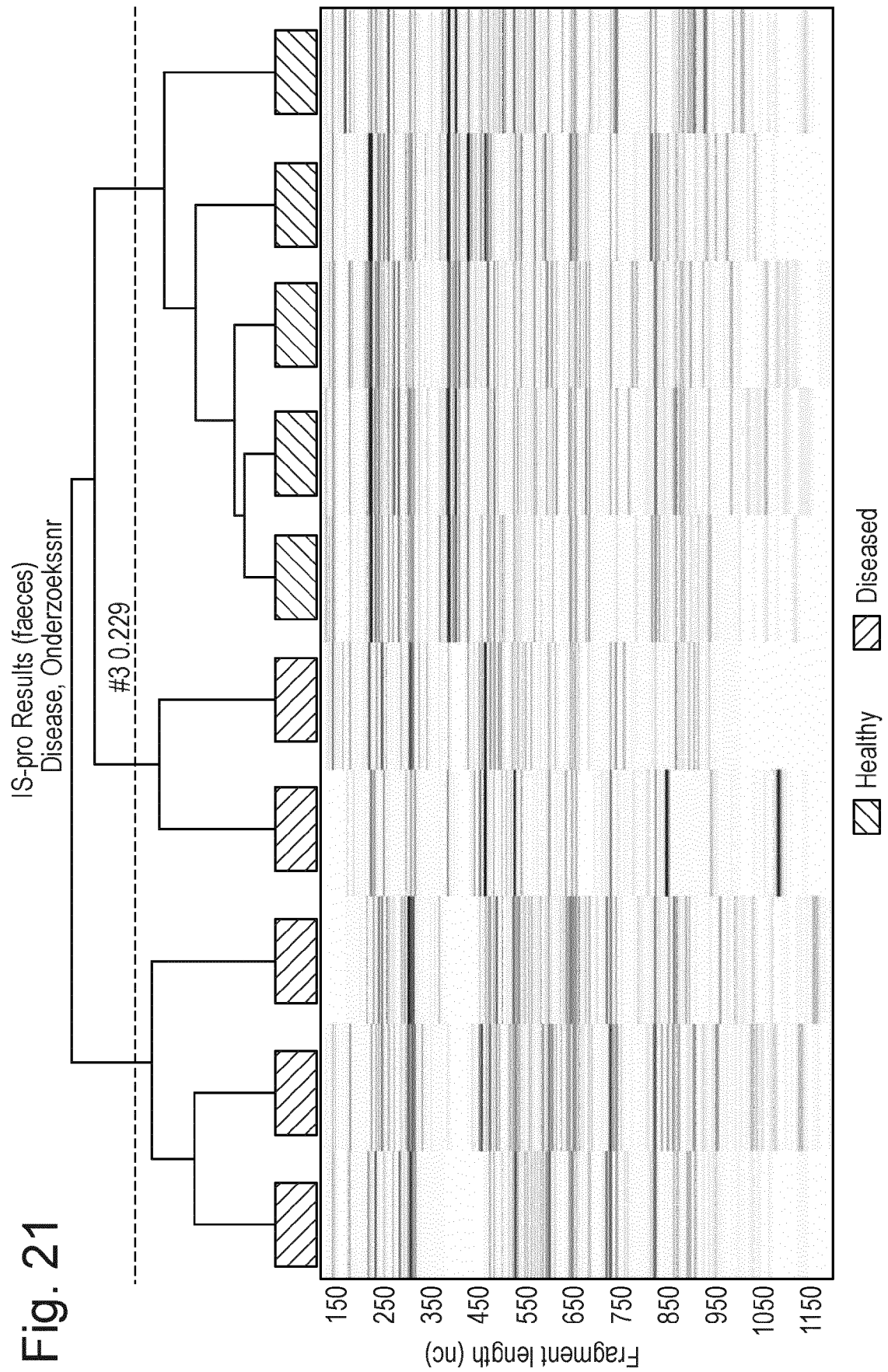


Fig. 19

22/55



23/55



24/55

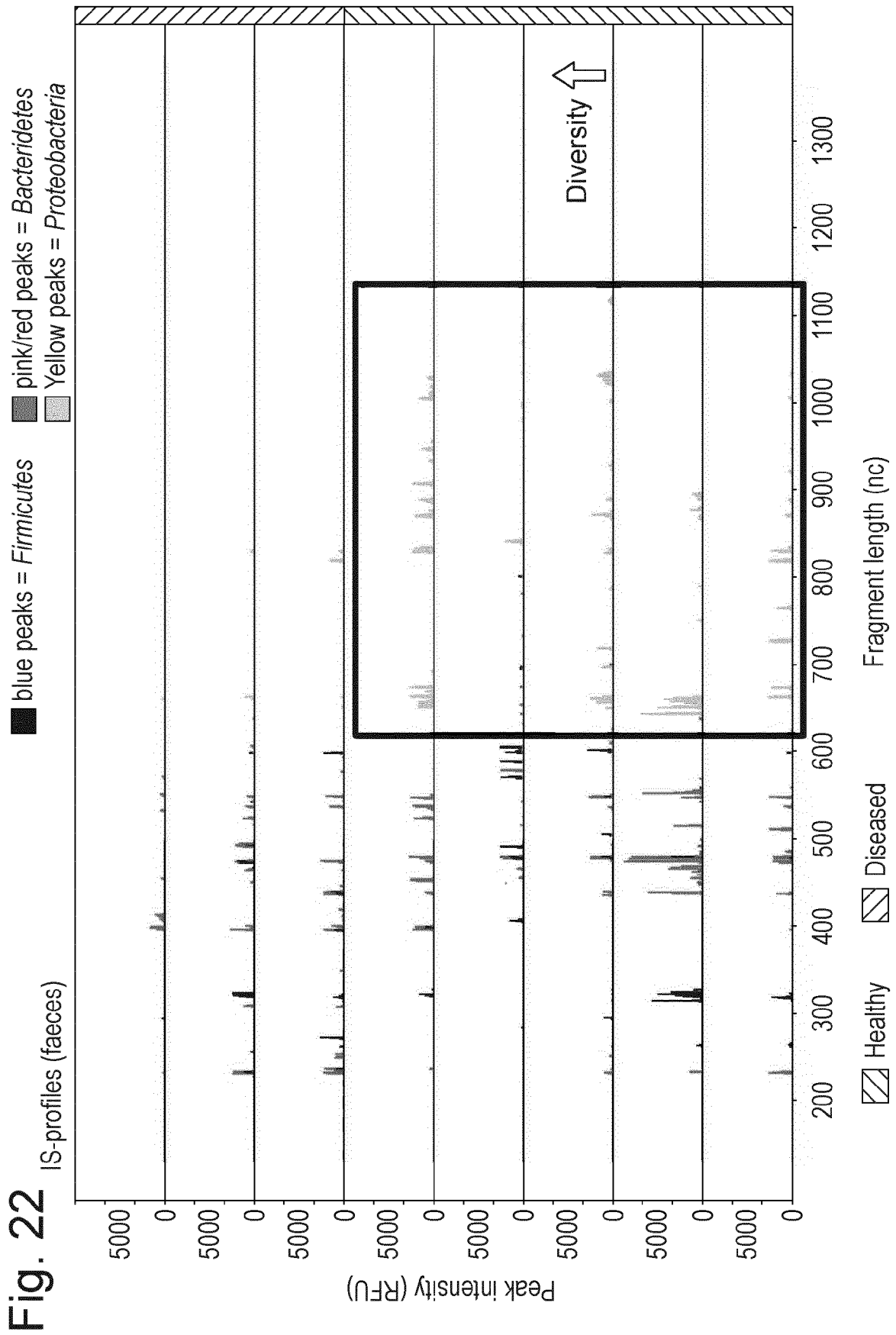
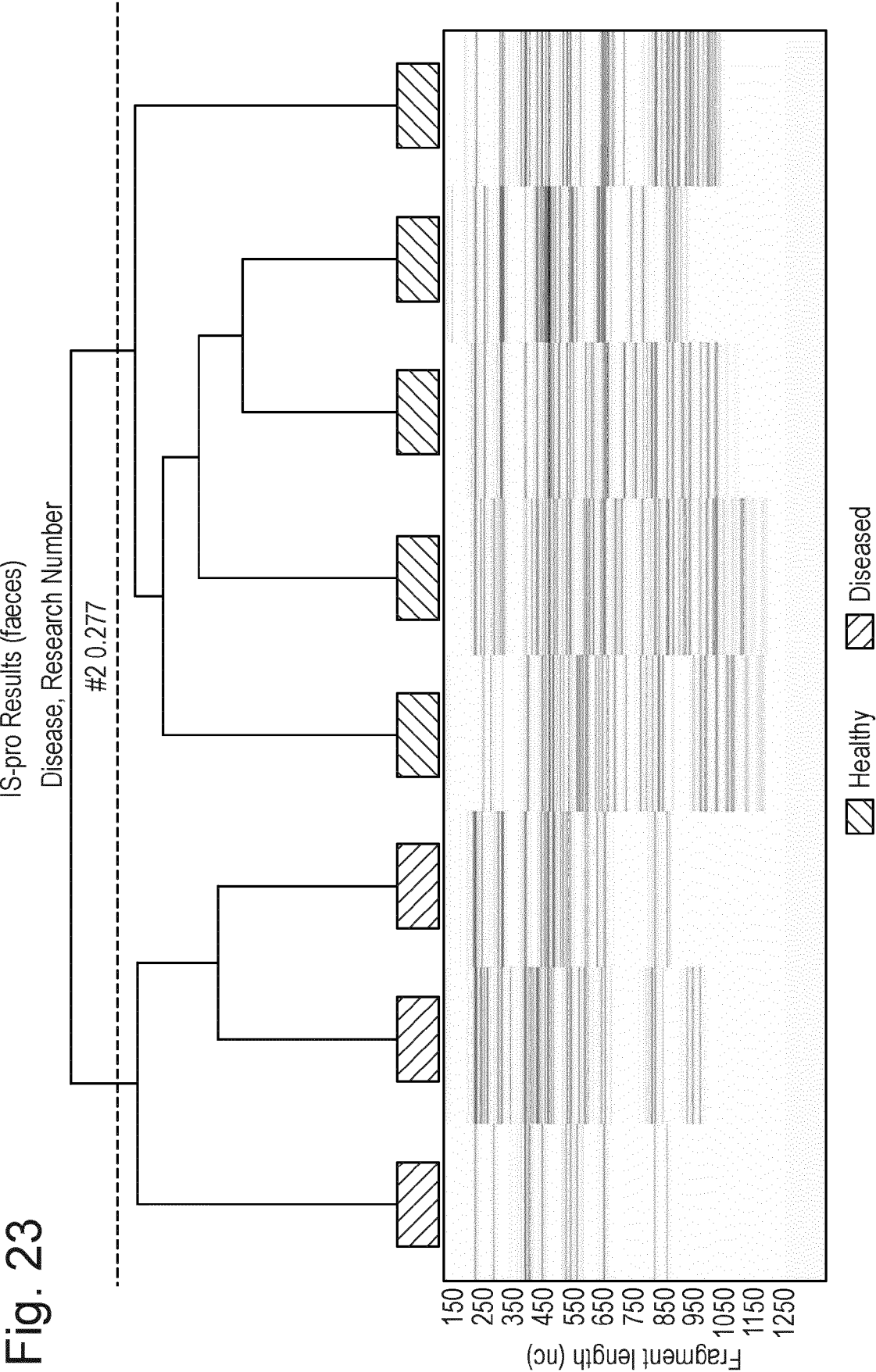
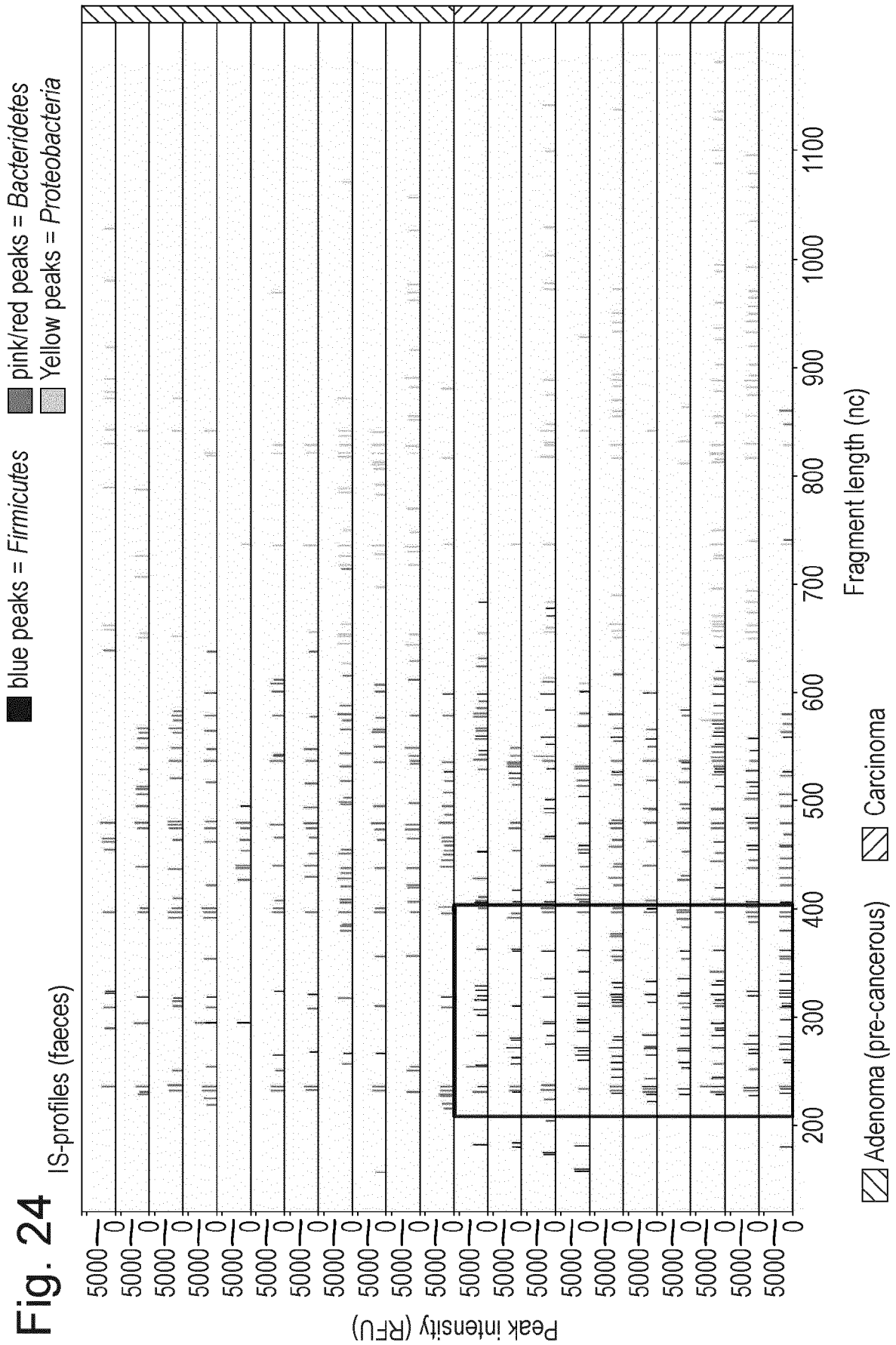
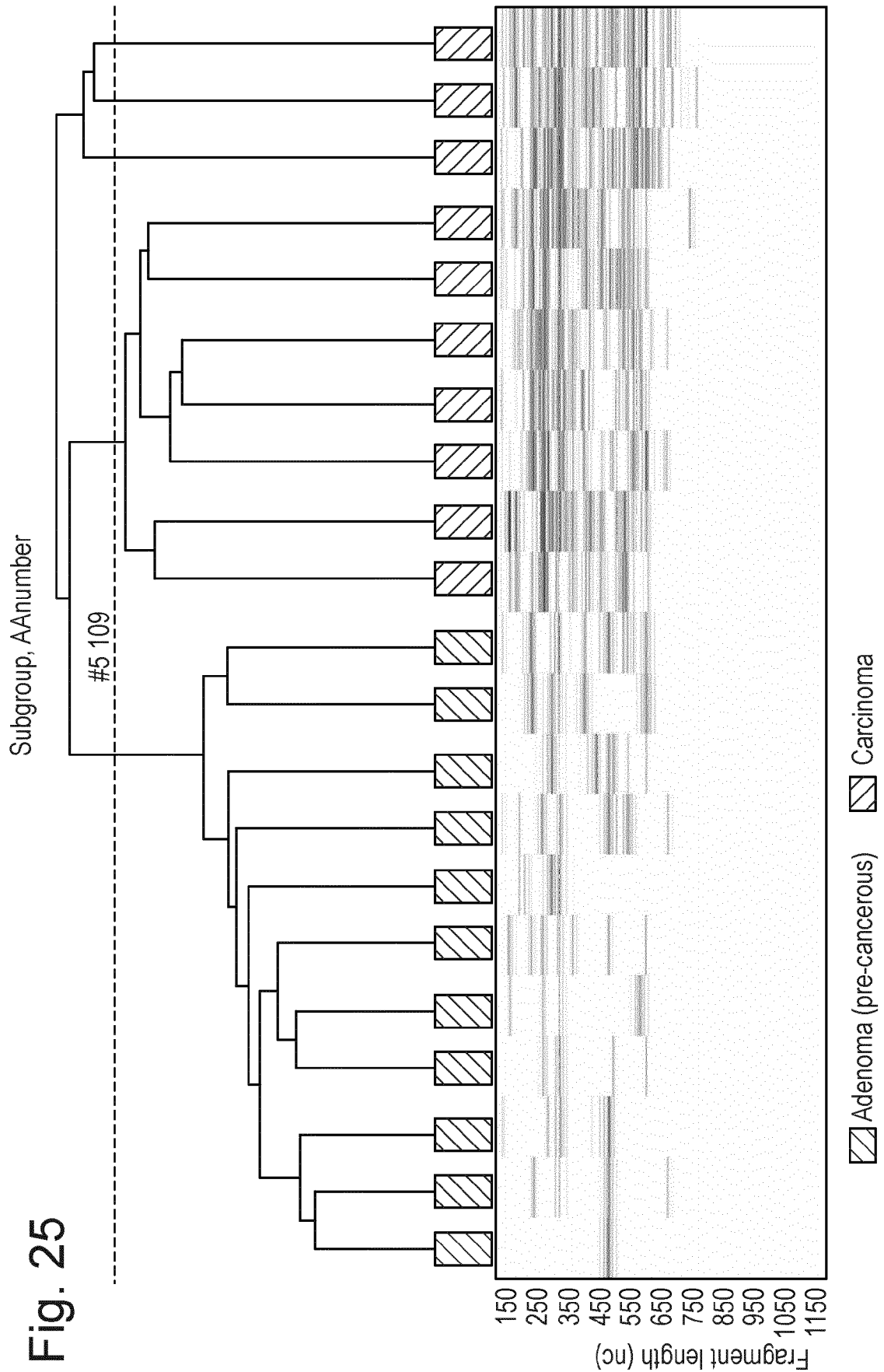


Fig. 23

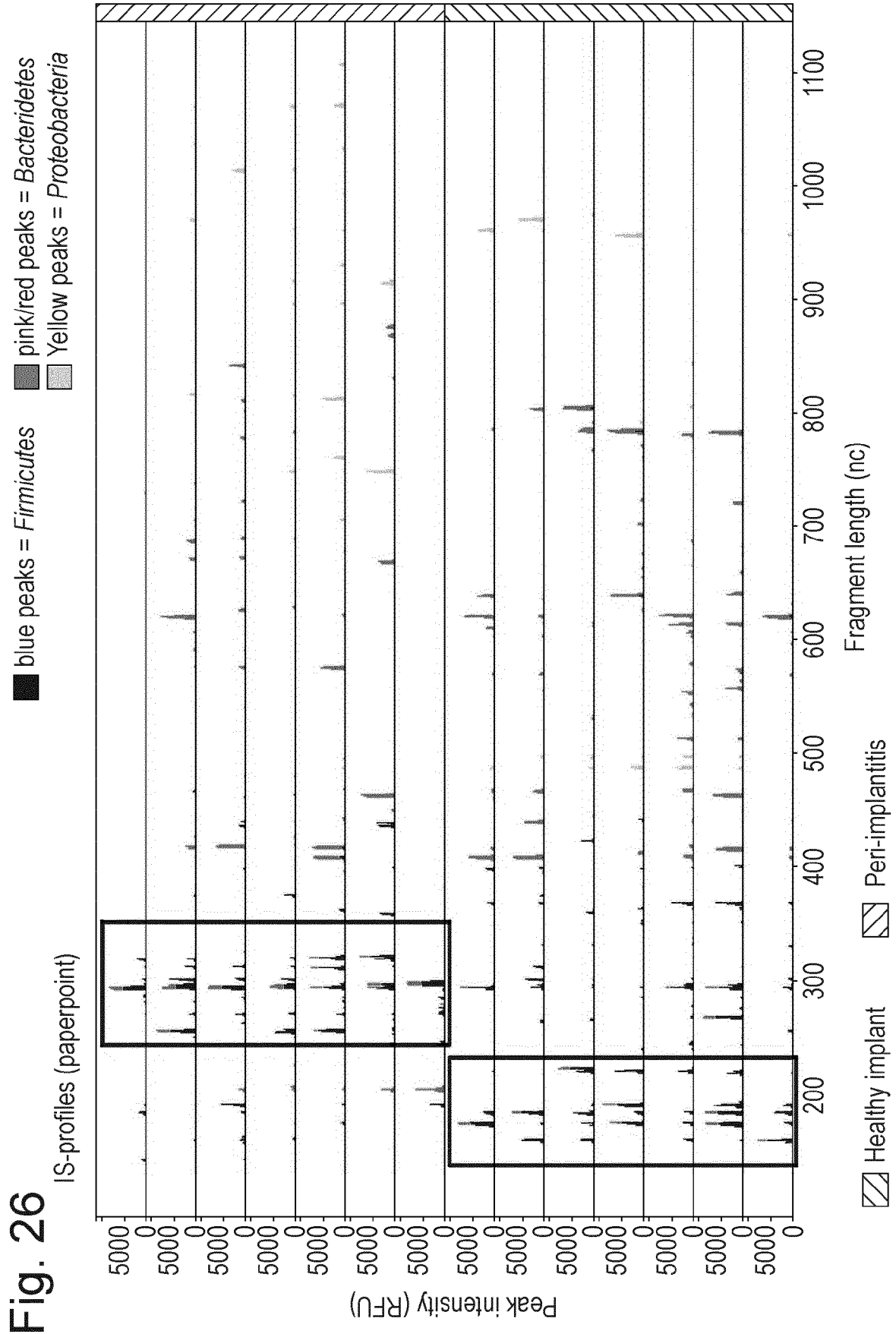




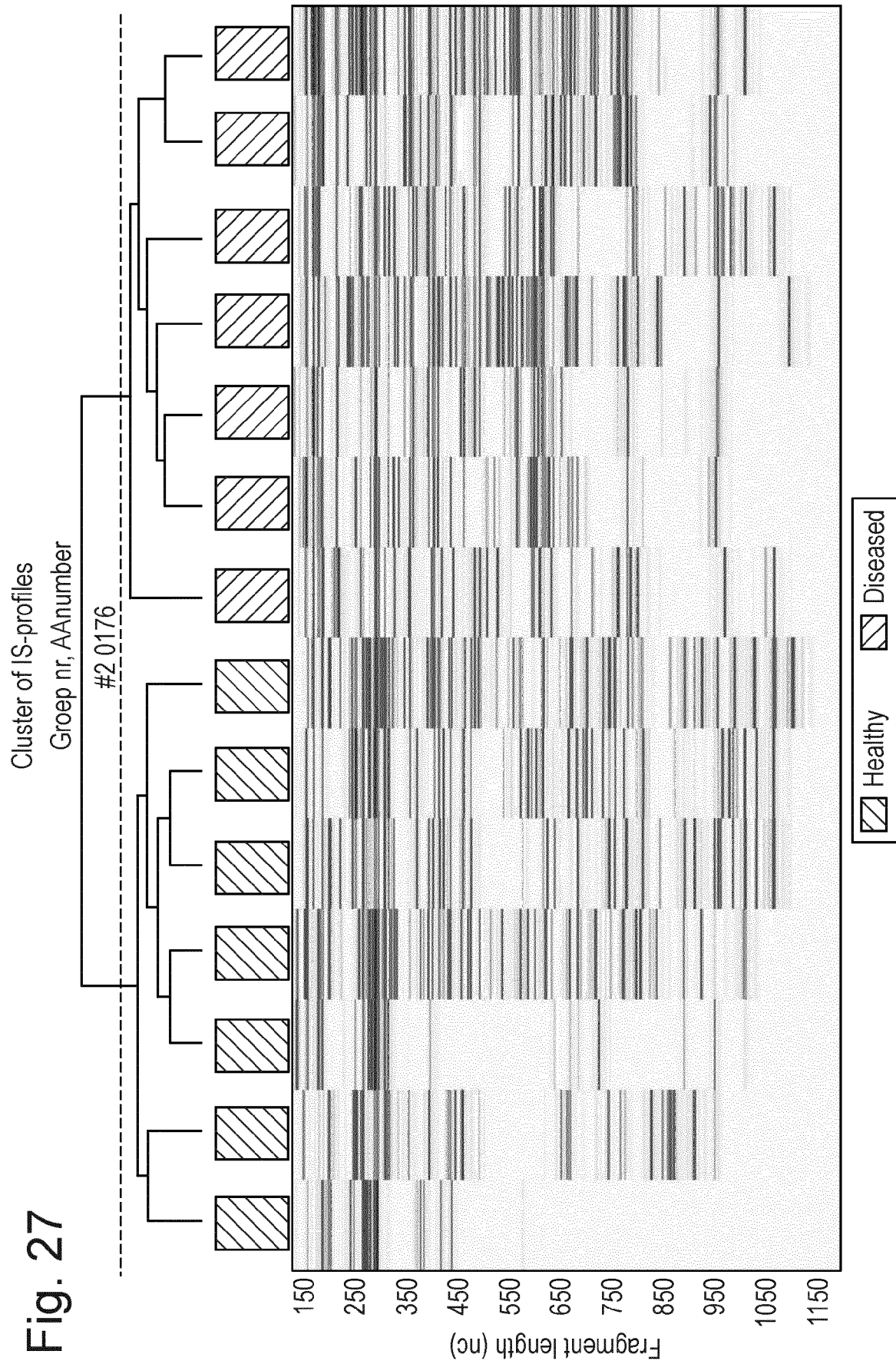
27/55



28/55



29/55



30/55

Fig. 28A

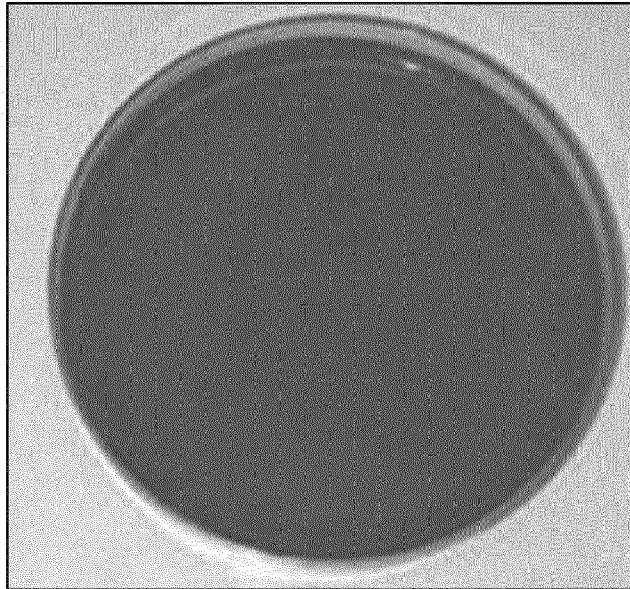


Fig. 28B

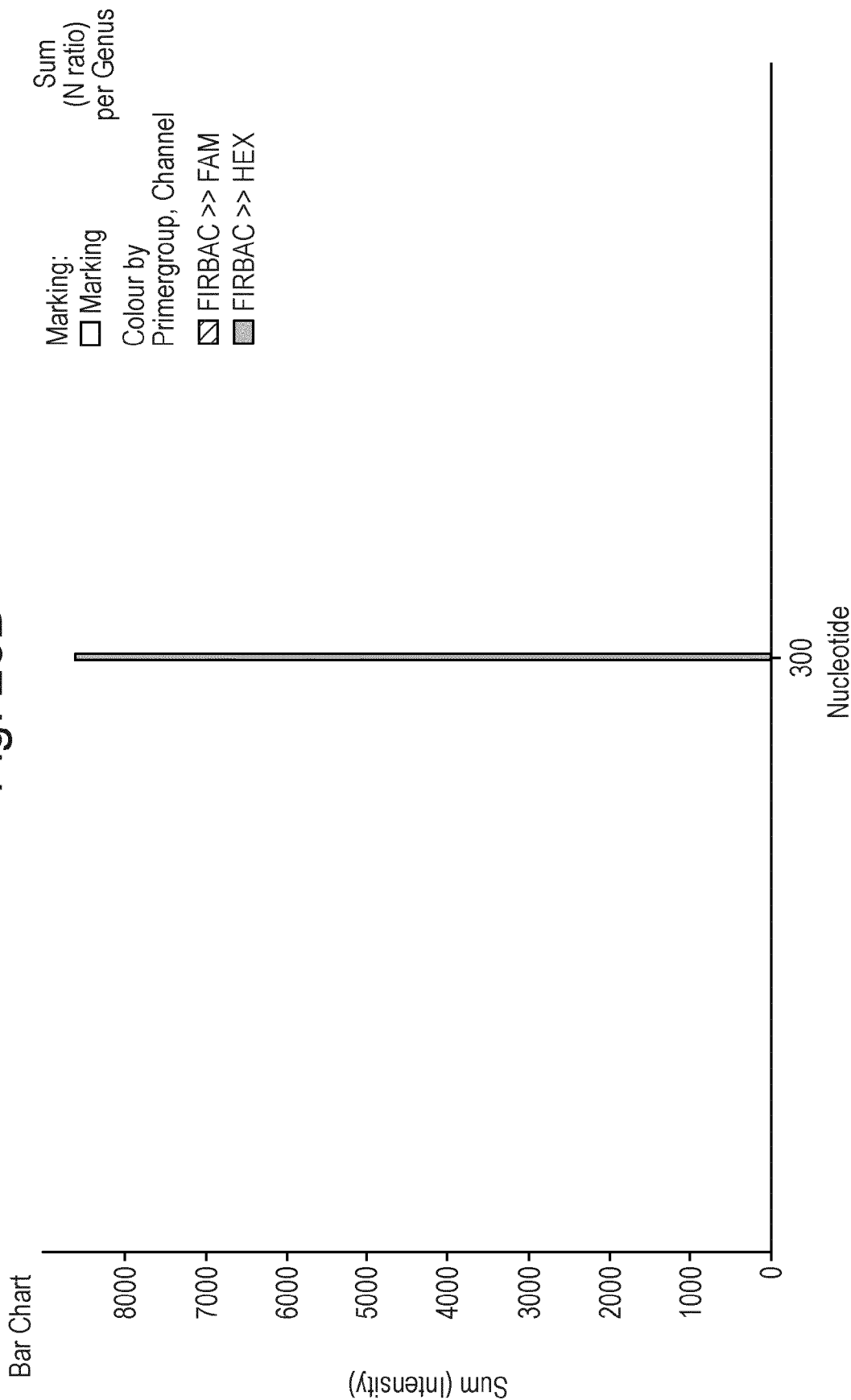
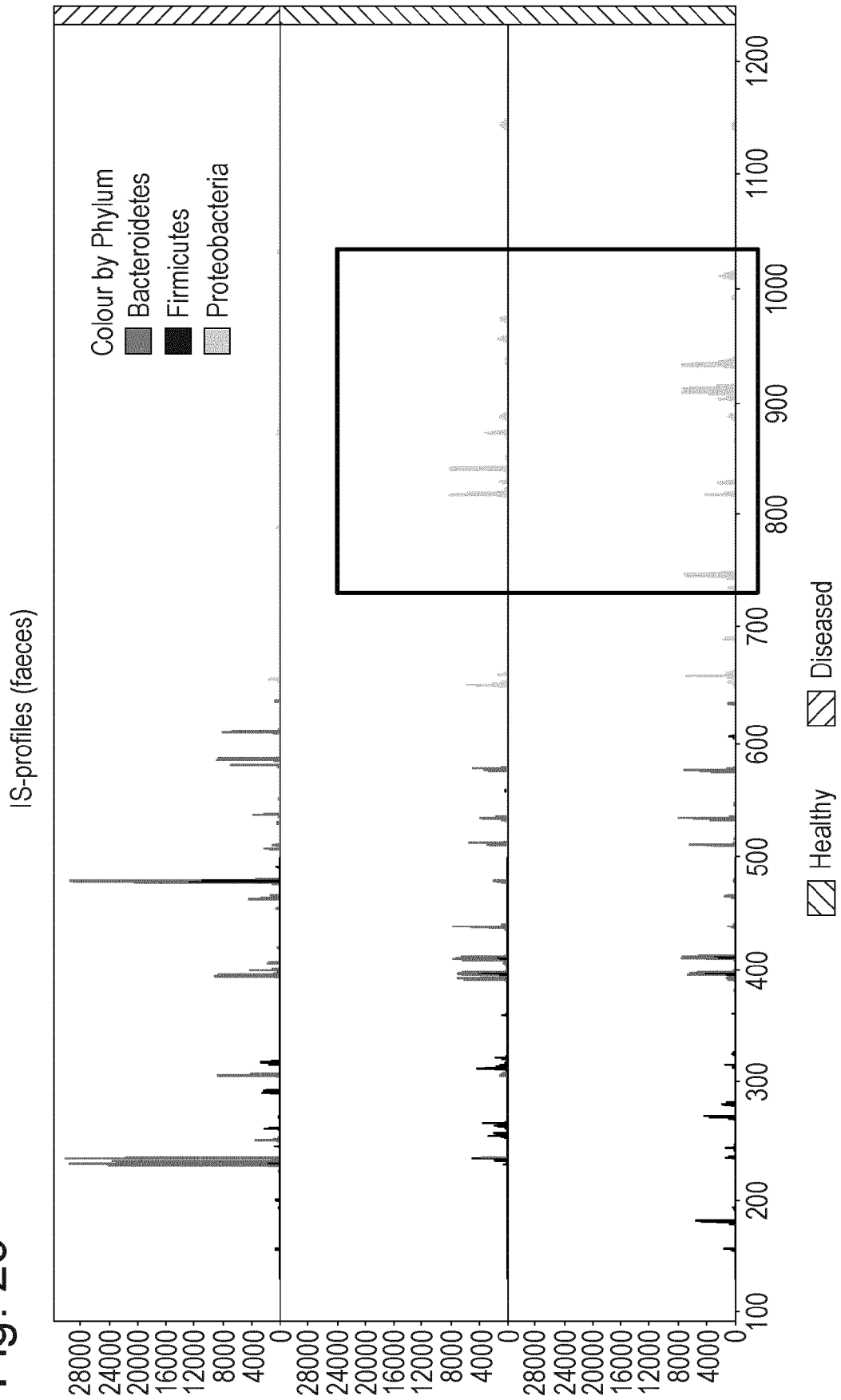


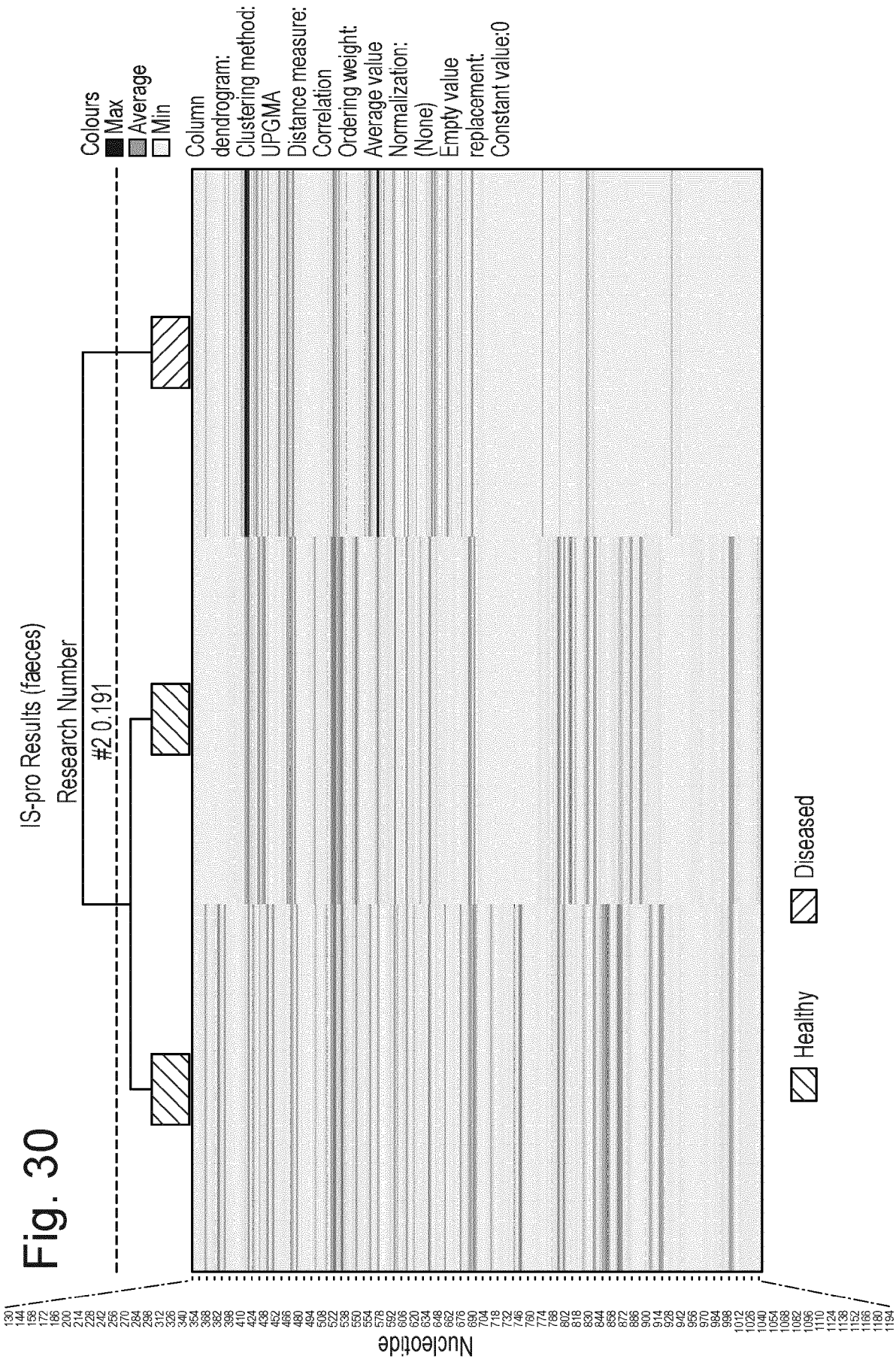
Fig. 28B (Cont.)

IS Lengtes 3							Data limited by: ☐ Marking
Source	N ratio	Organism	Genus	Species	Kingdom	Phylum	
HMP	0,00	AEVC01000028...	Streptococcus	cristatus	Bacteria	Firmicutes	Data table: IS Lengtes 3 Marking: (None)
HMP	0,00	GL732519 Strept...	Streptococcus	cristatus	Bacteria	Firmicutes	

33/55

Fig. 29





35/55

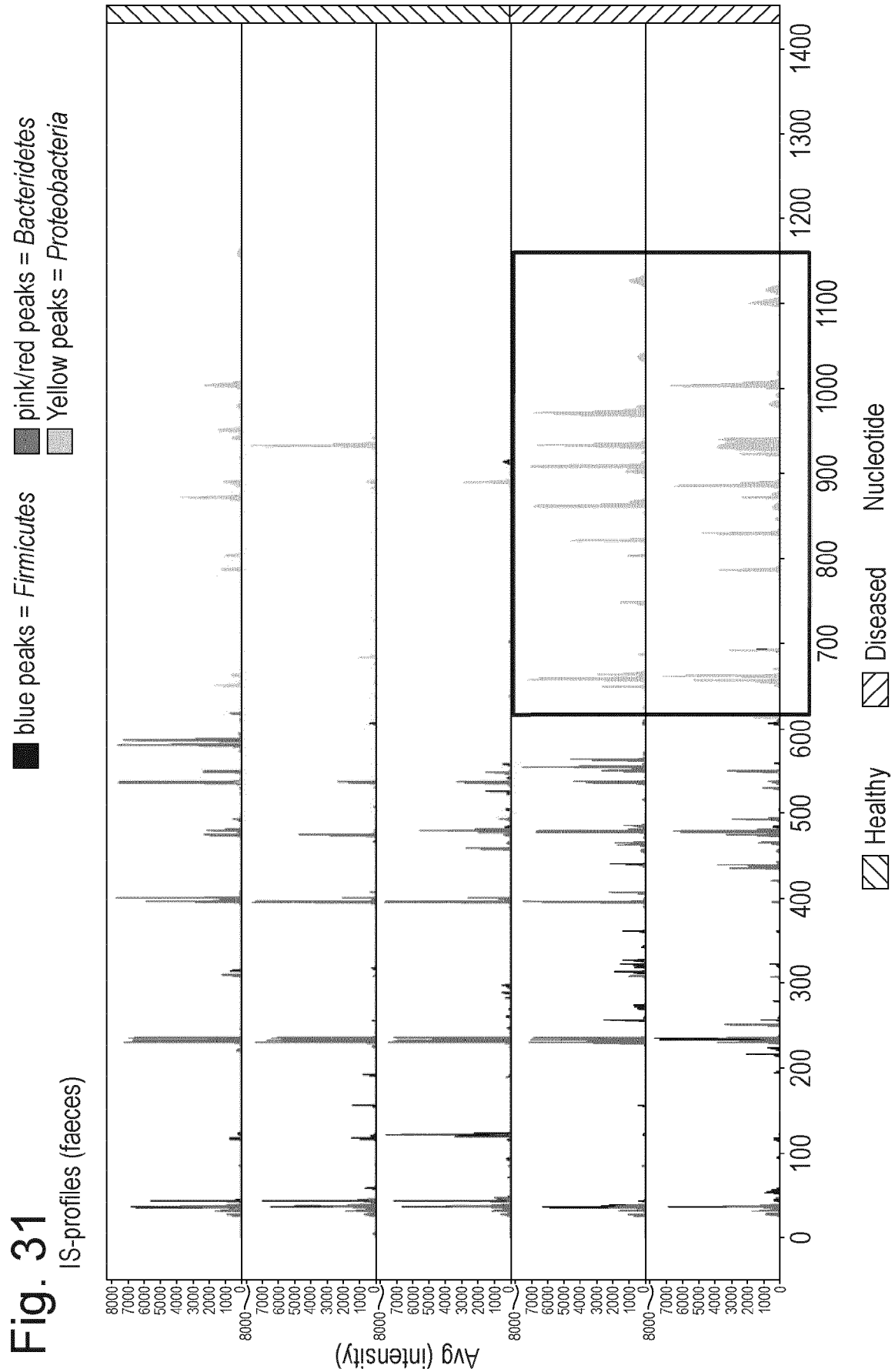
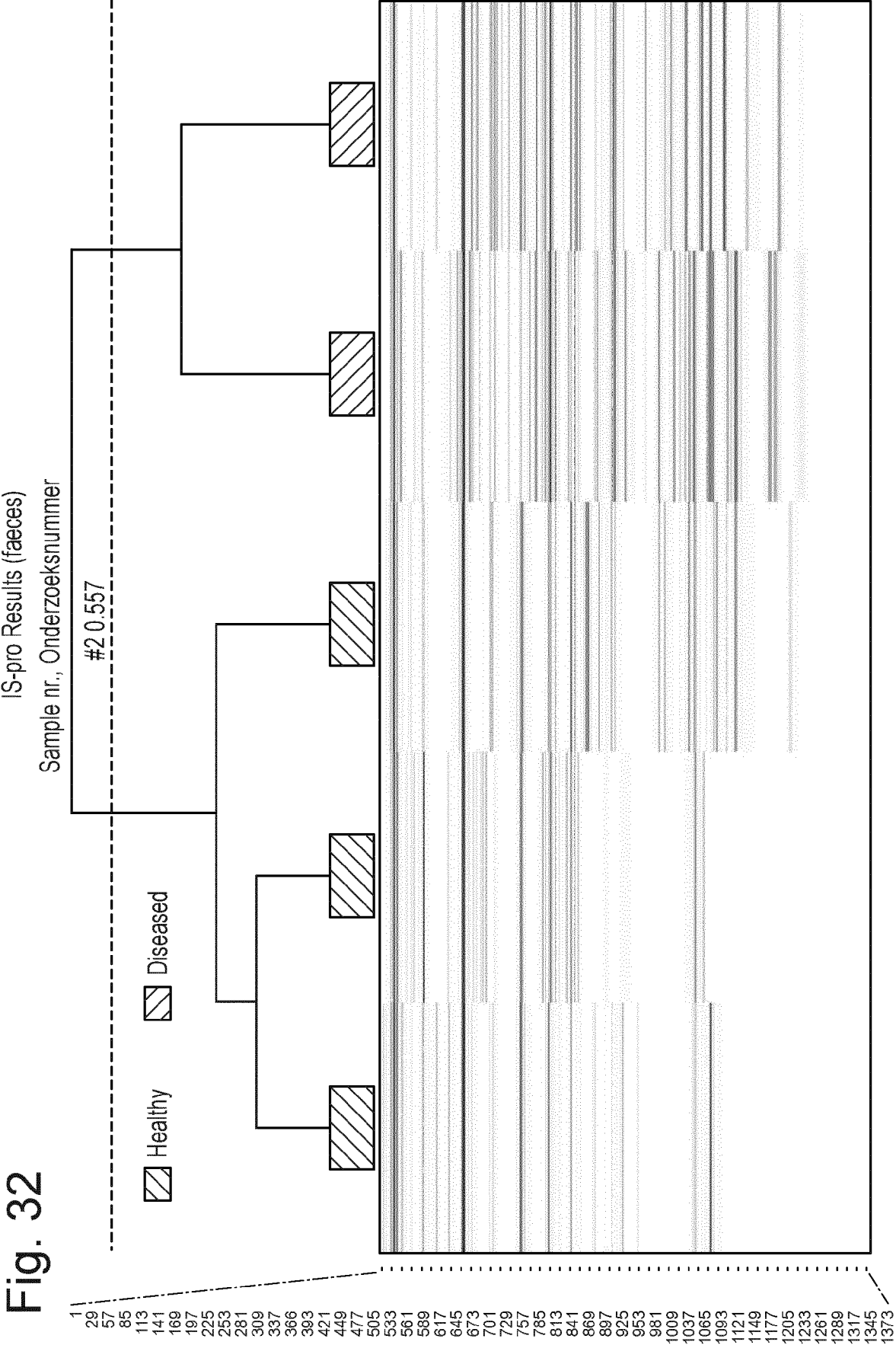
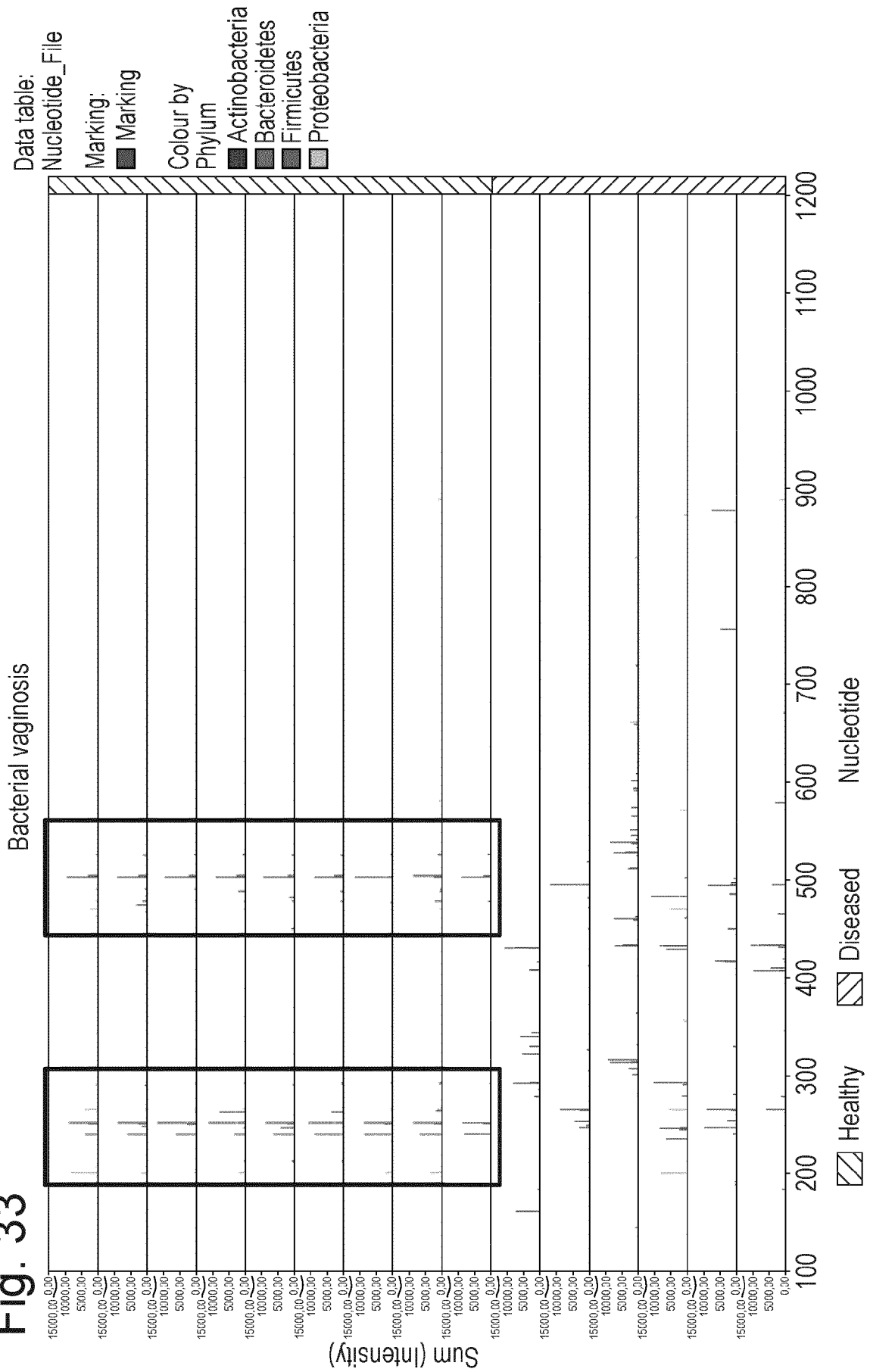


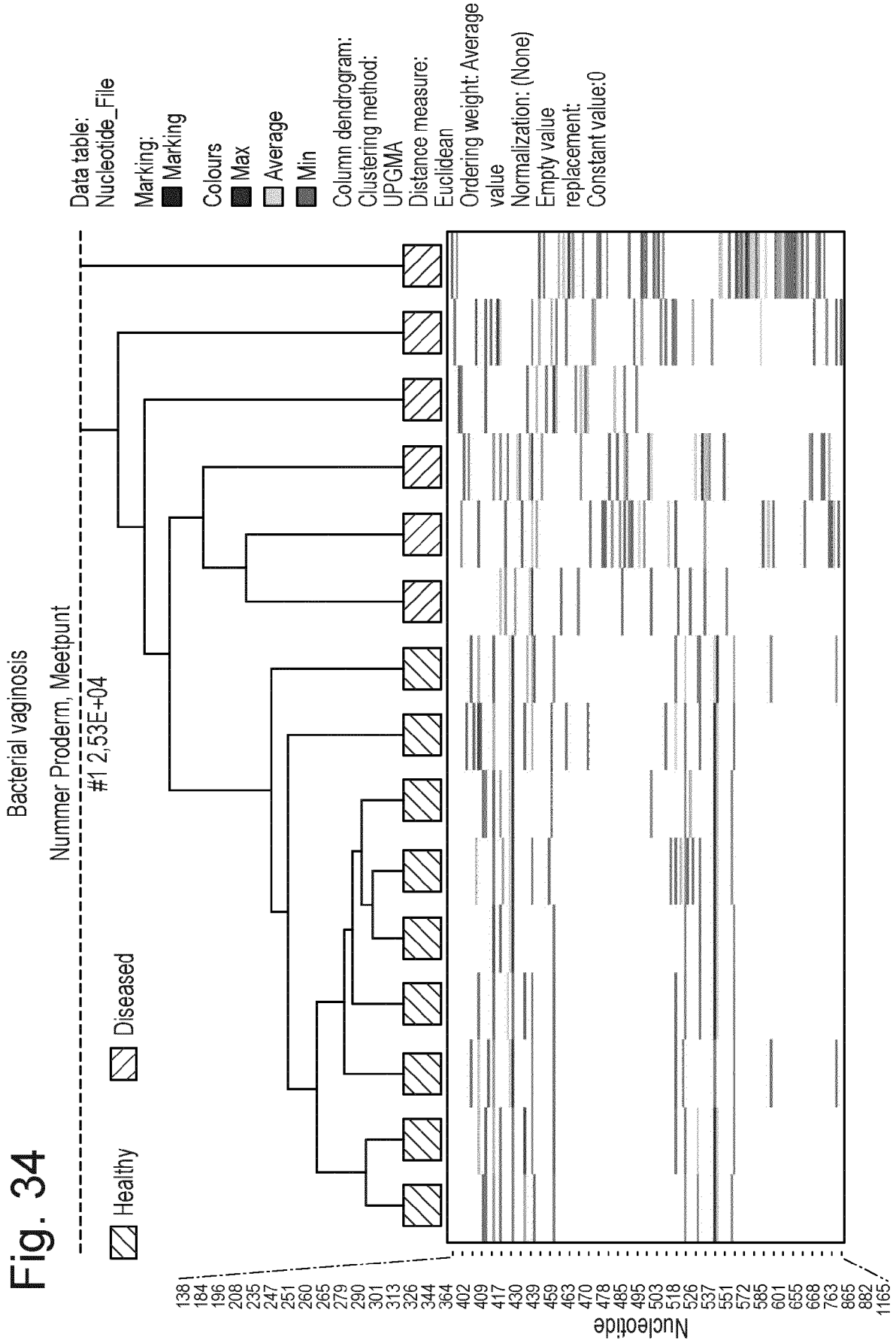
Fig. 32



37/55

Fig. 33





39/55

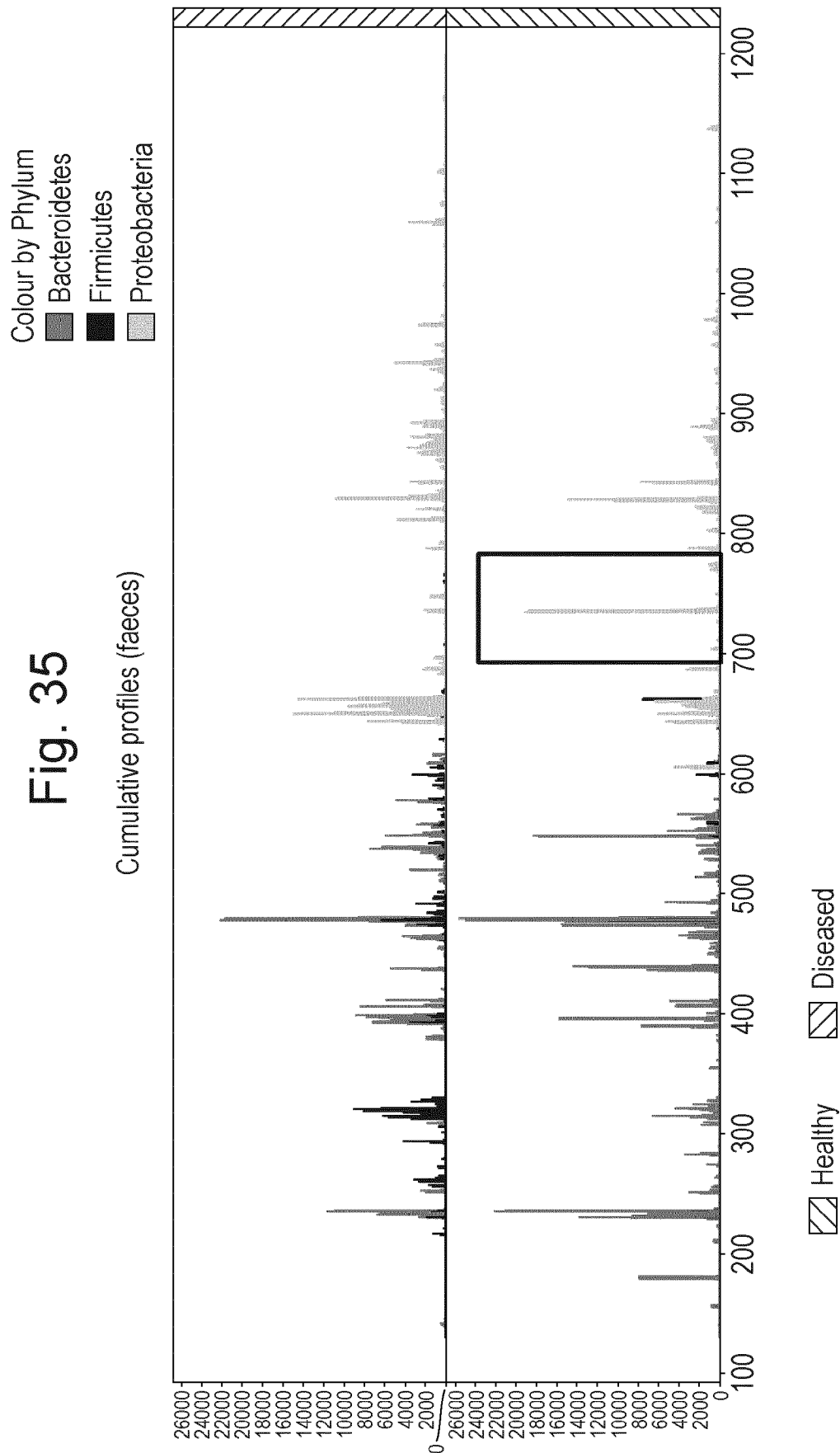
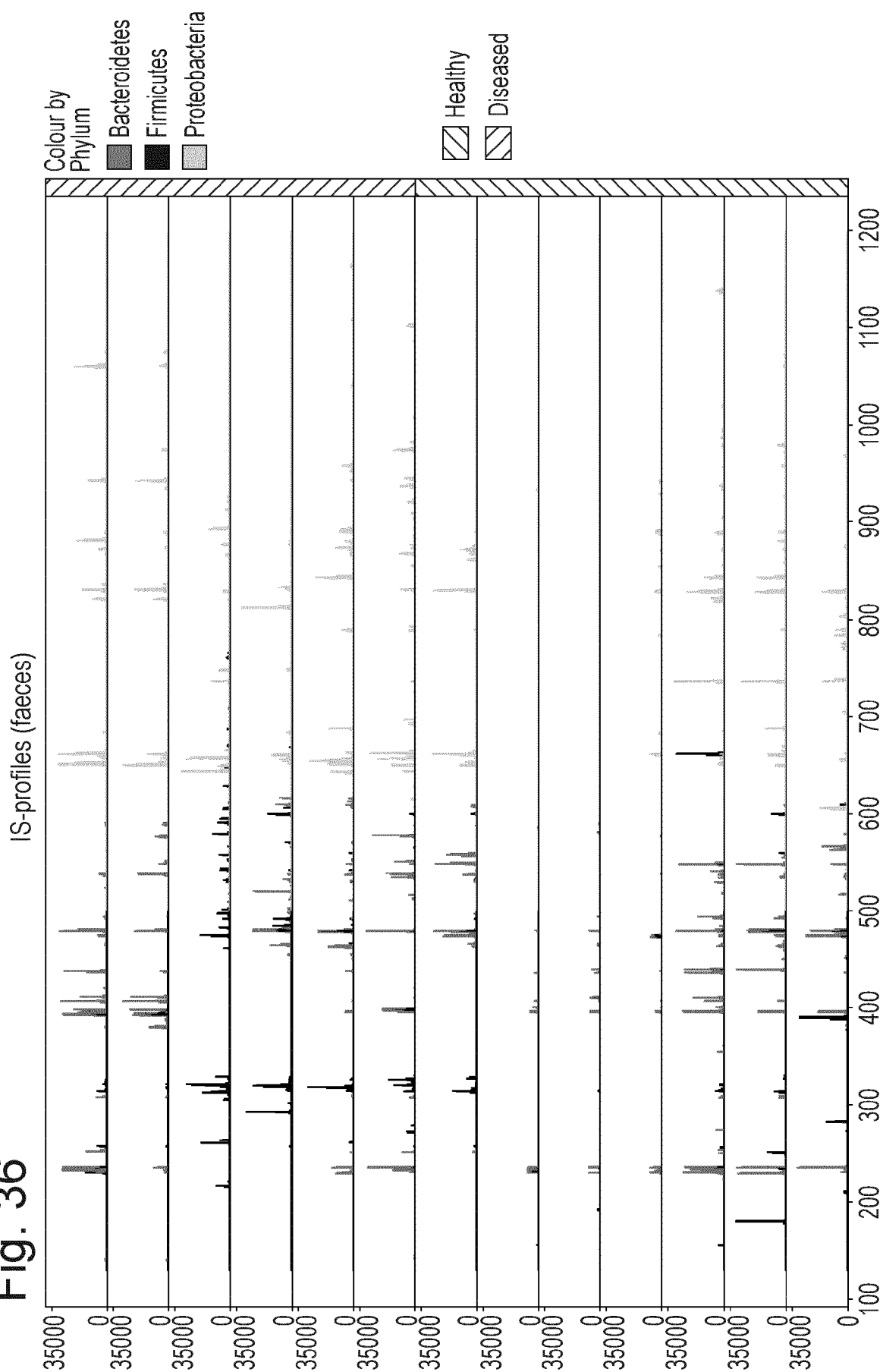
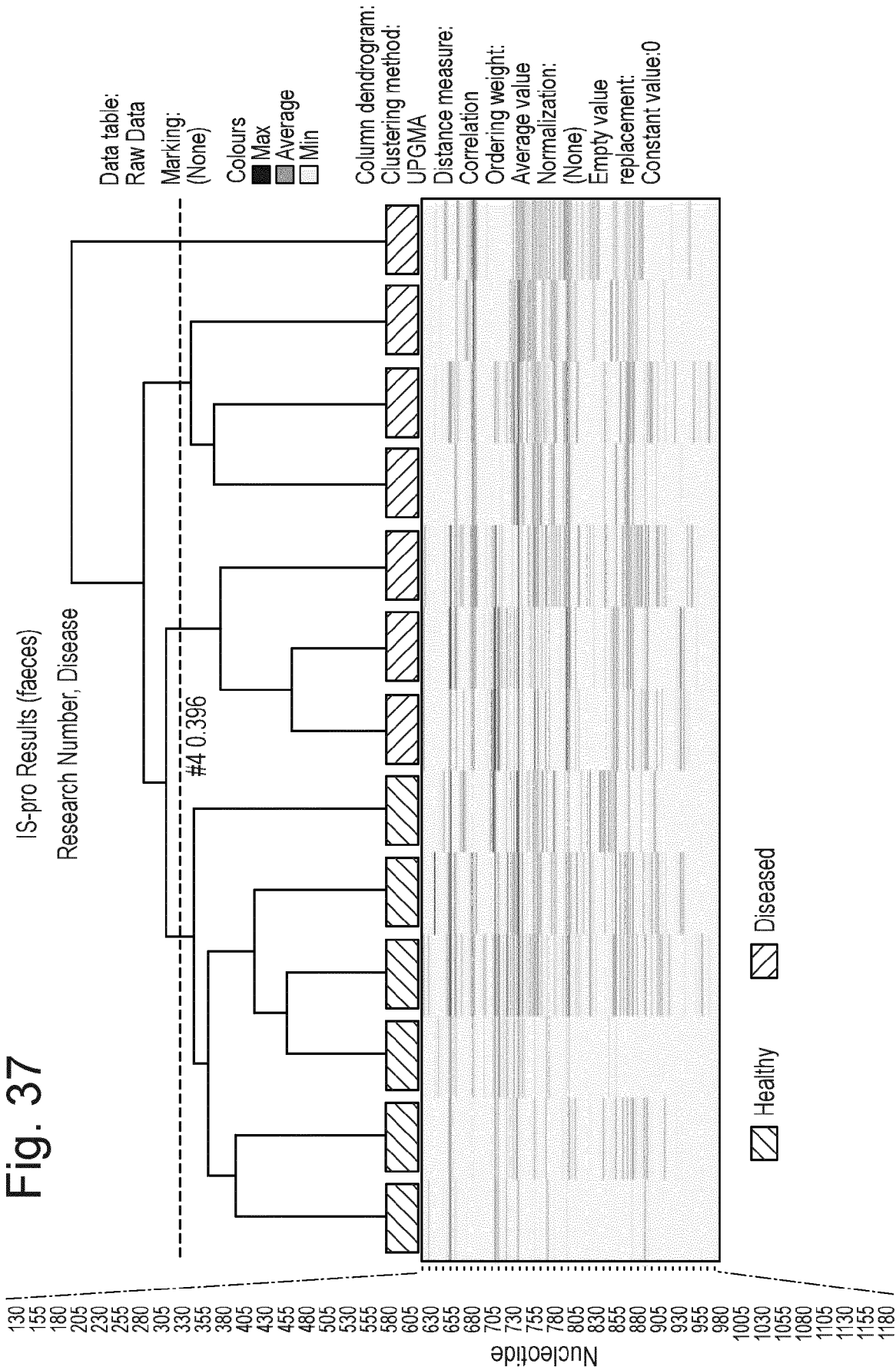


Fig. 36



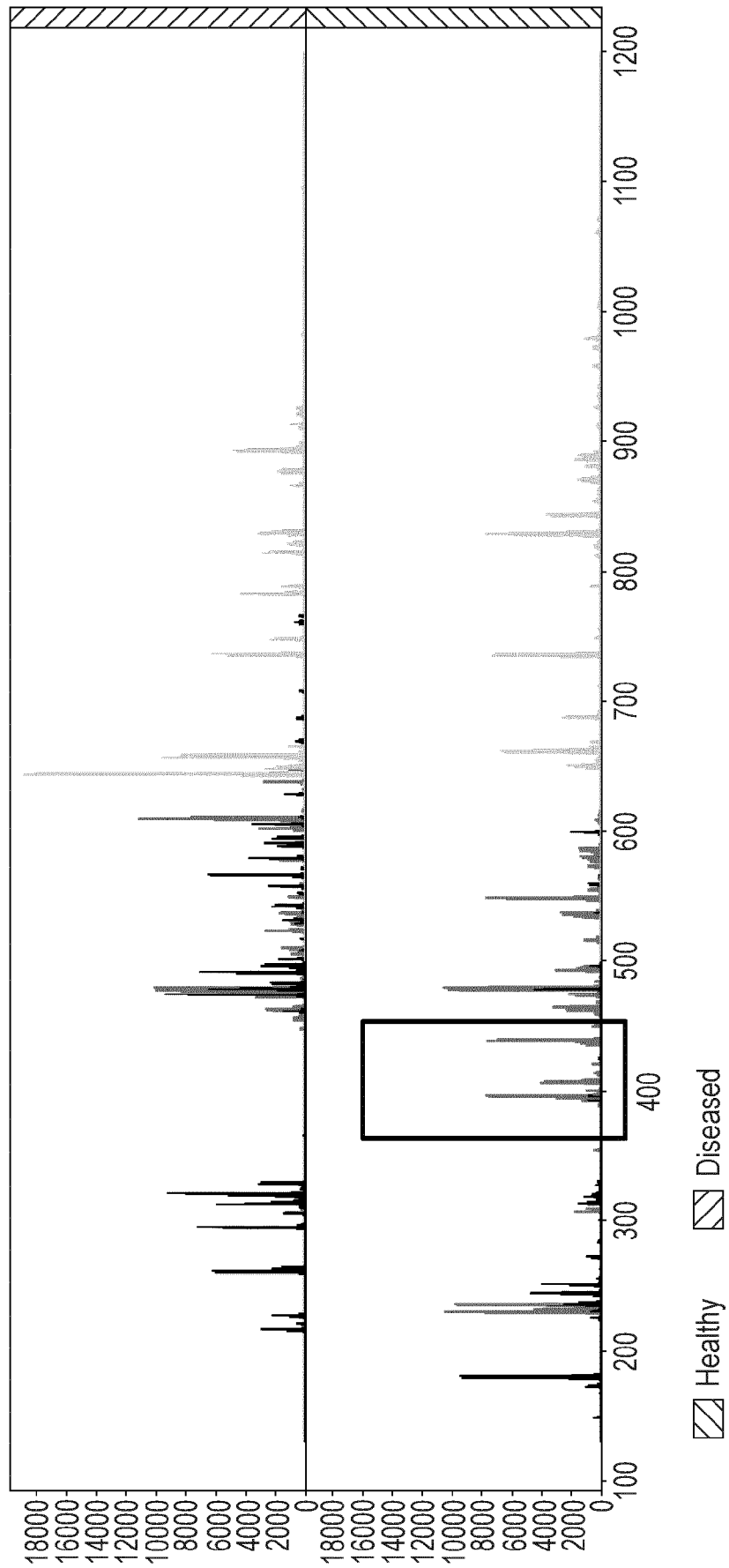


42/55

Colour by Phylum
Bacteroidetes
Firmicutes
Proteobacteria

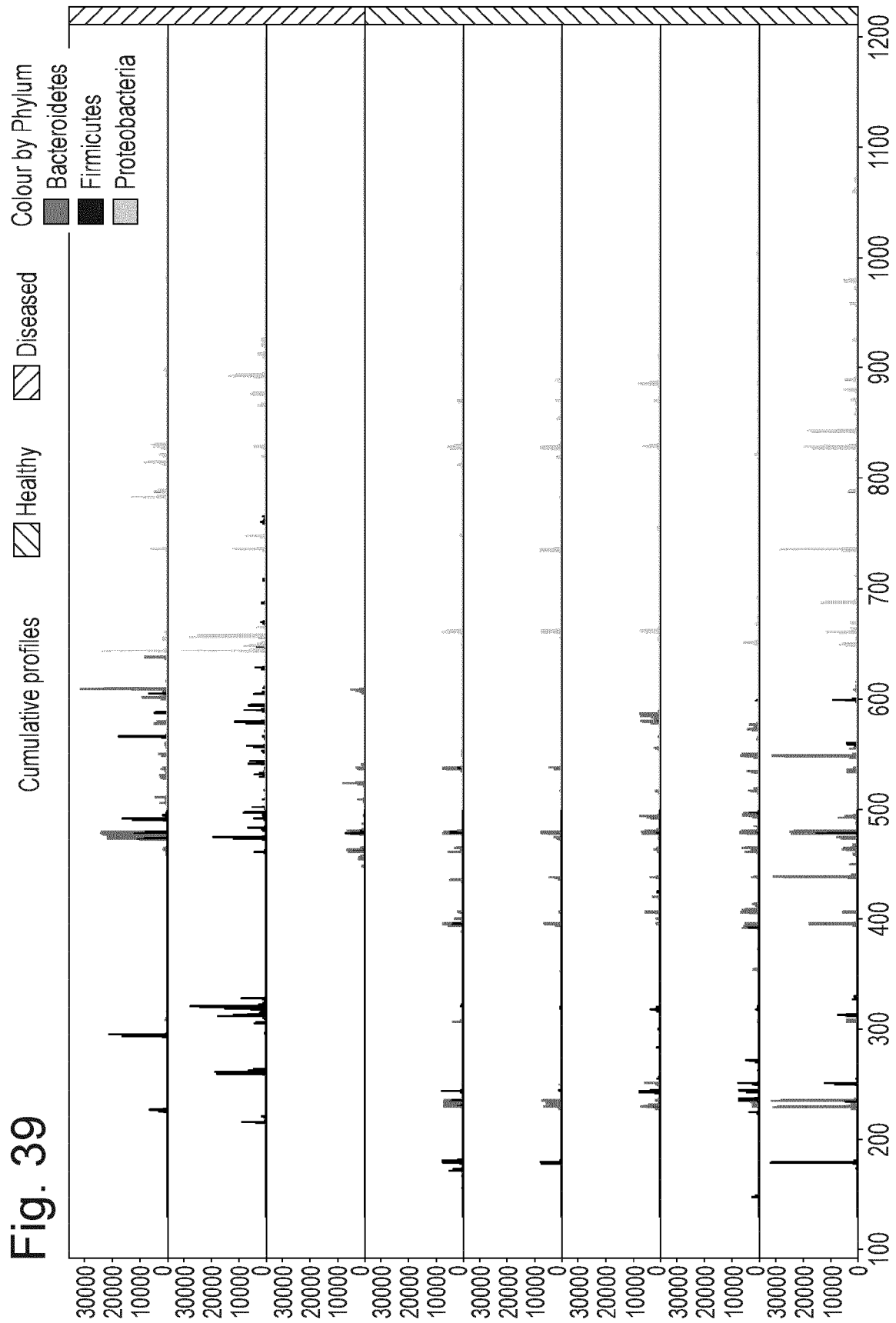
Fig. 38

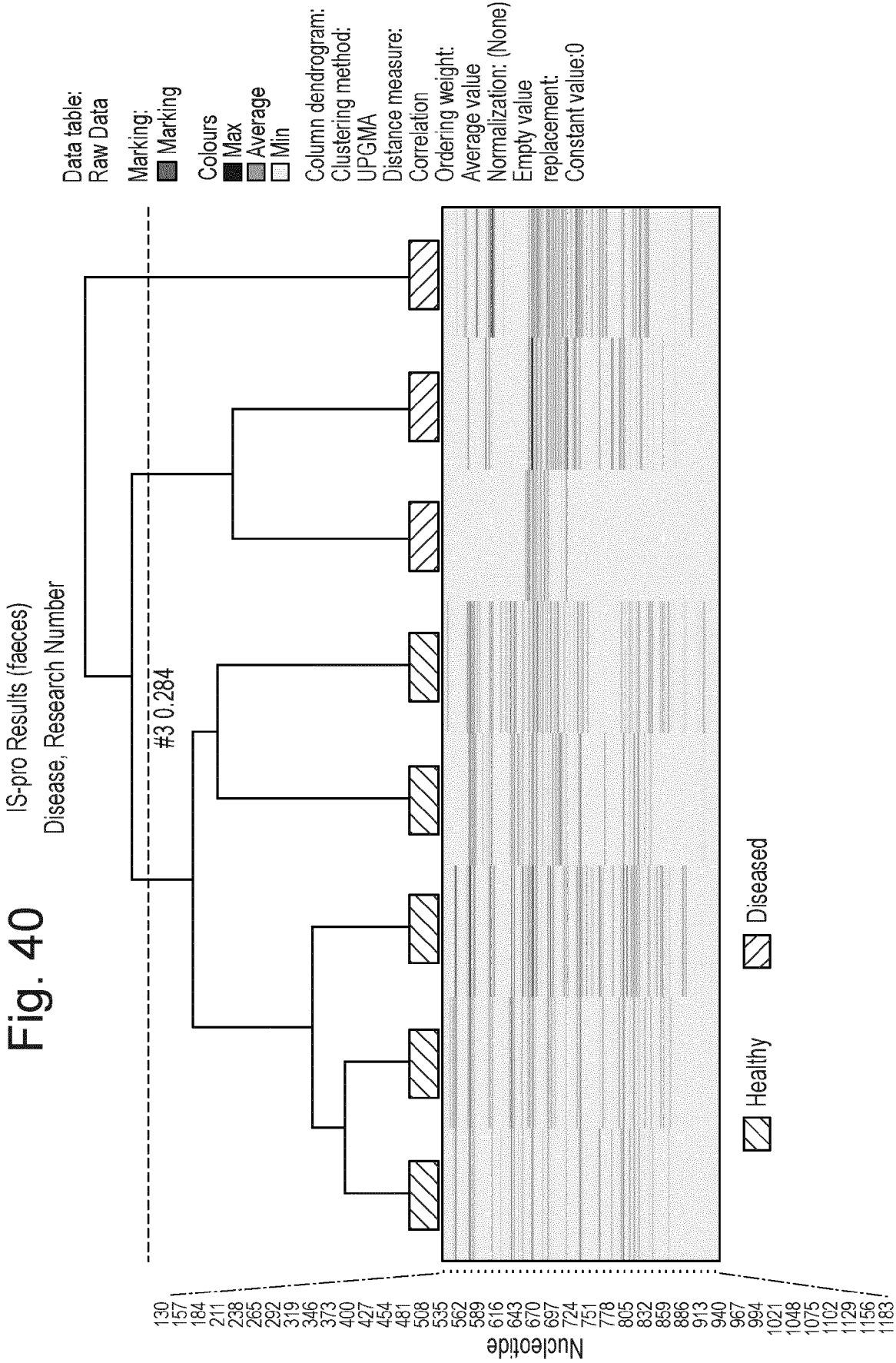
Cumulative profiles (faeces)



SUBSTITUTE SHEET (RULE 26)

43/55





45/55

Fig. 41A

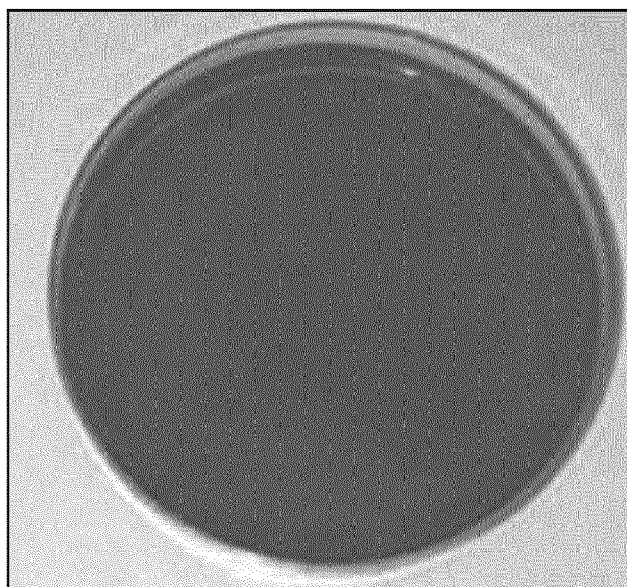
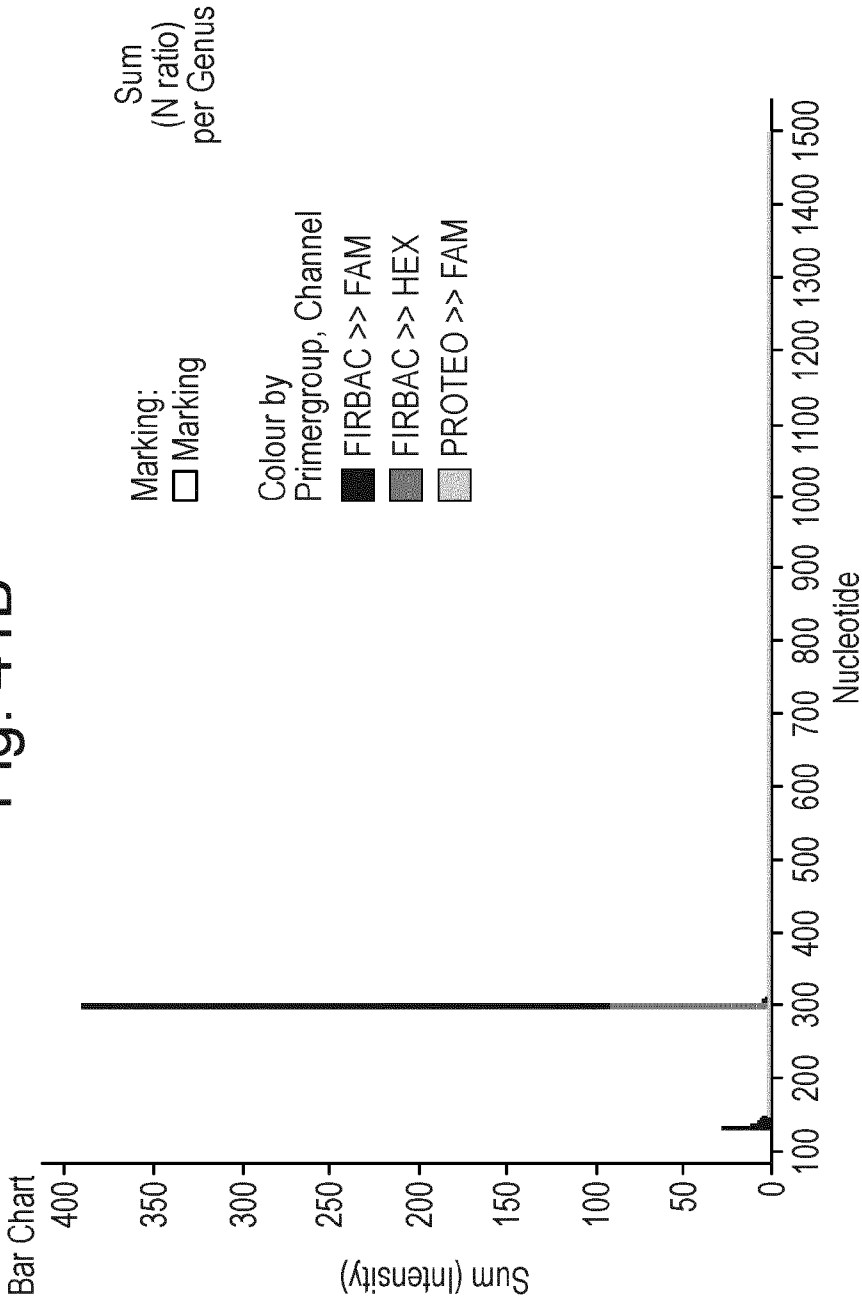


Fig. 41B



IS Lengtes 3

Source	N ratio	Organism	Genus	Species	Kingdom	Phylum	Class
NCBI	0,00	Streptococcus m...	Streptococcus	mitis	Bacteria	Firmicutes	Bacili
NCBI	0,00	Streptococcus m...	Streptococcus	mitis	Bacteria	Firmicutes	Bacili
NCBI	0,00	Streptococcus m...	Streptococcus	mitis	Bacteria	Firmicutes	Bacili

Data limited by:

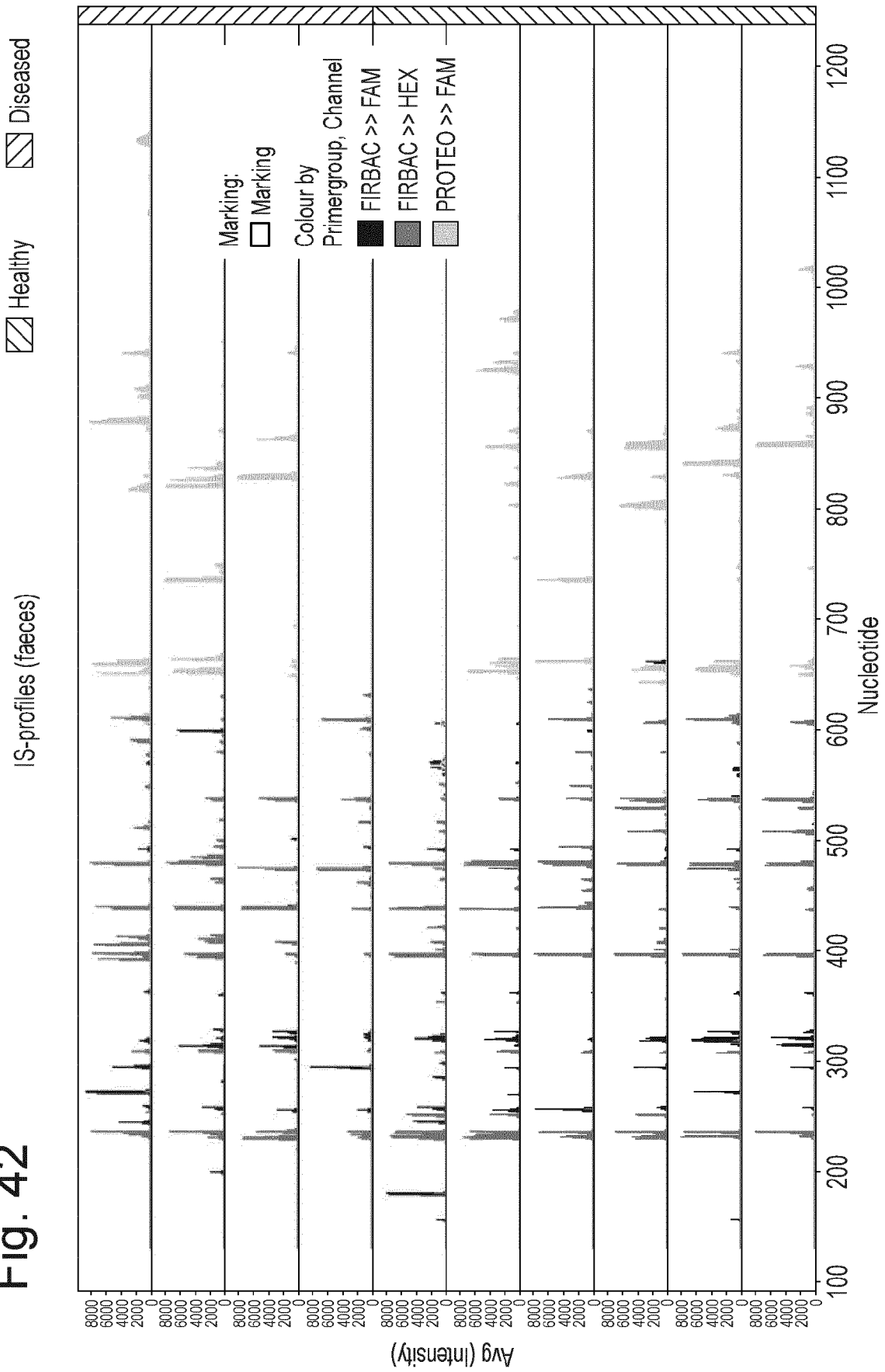
Data table:

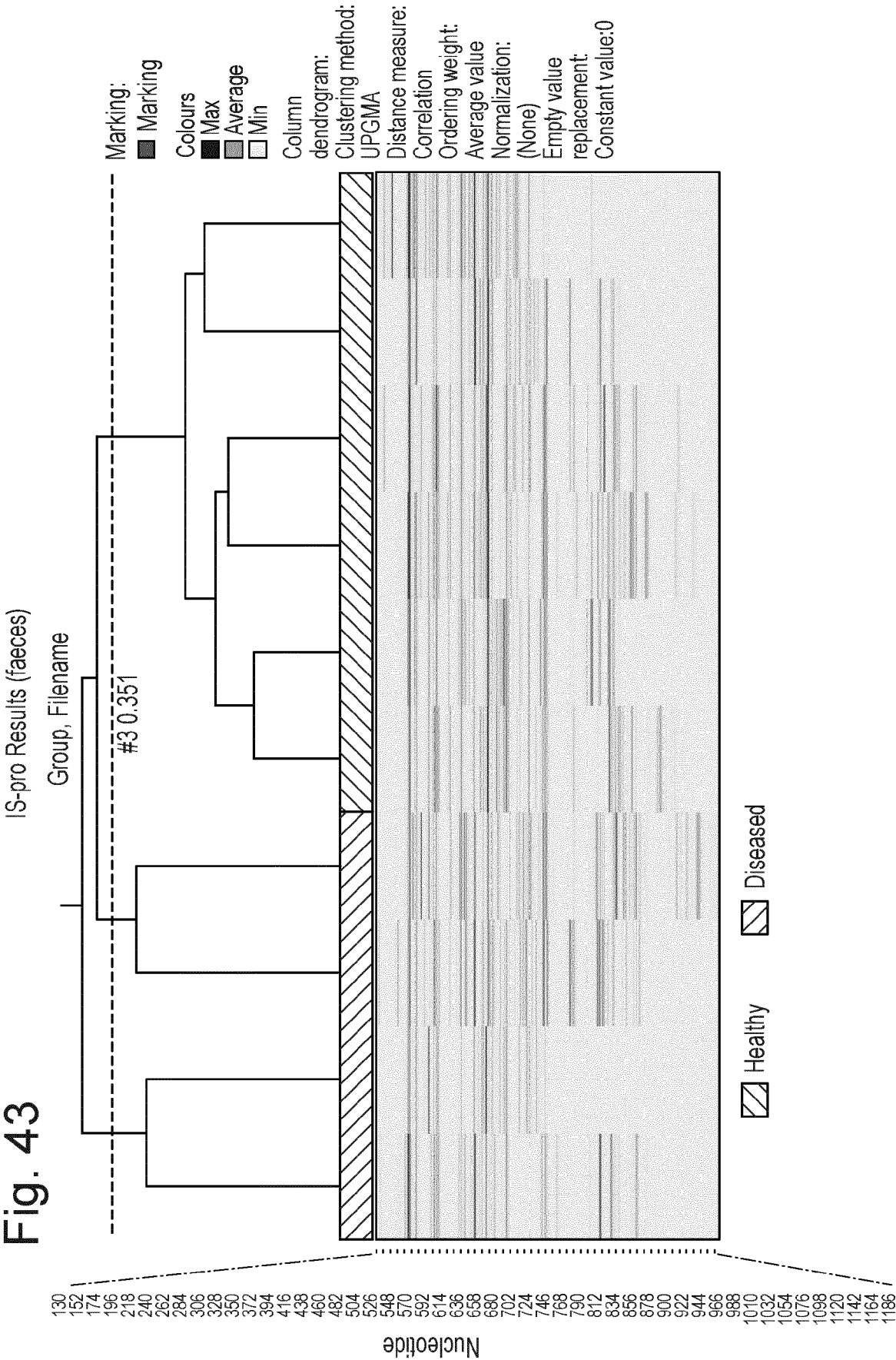
IS Lengtes 3

Marking:

47/55

Fig. 42





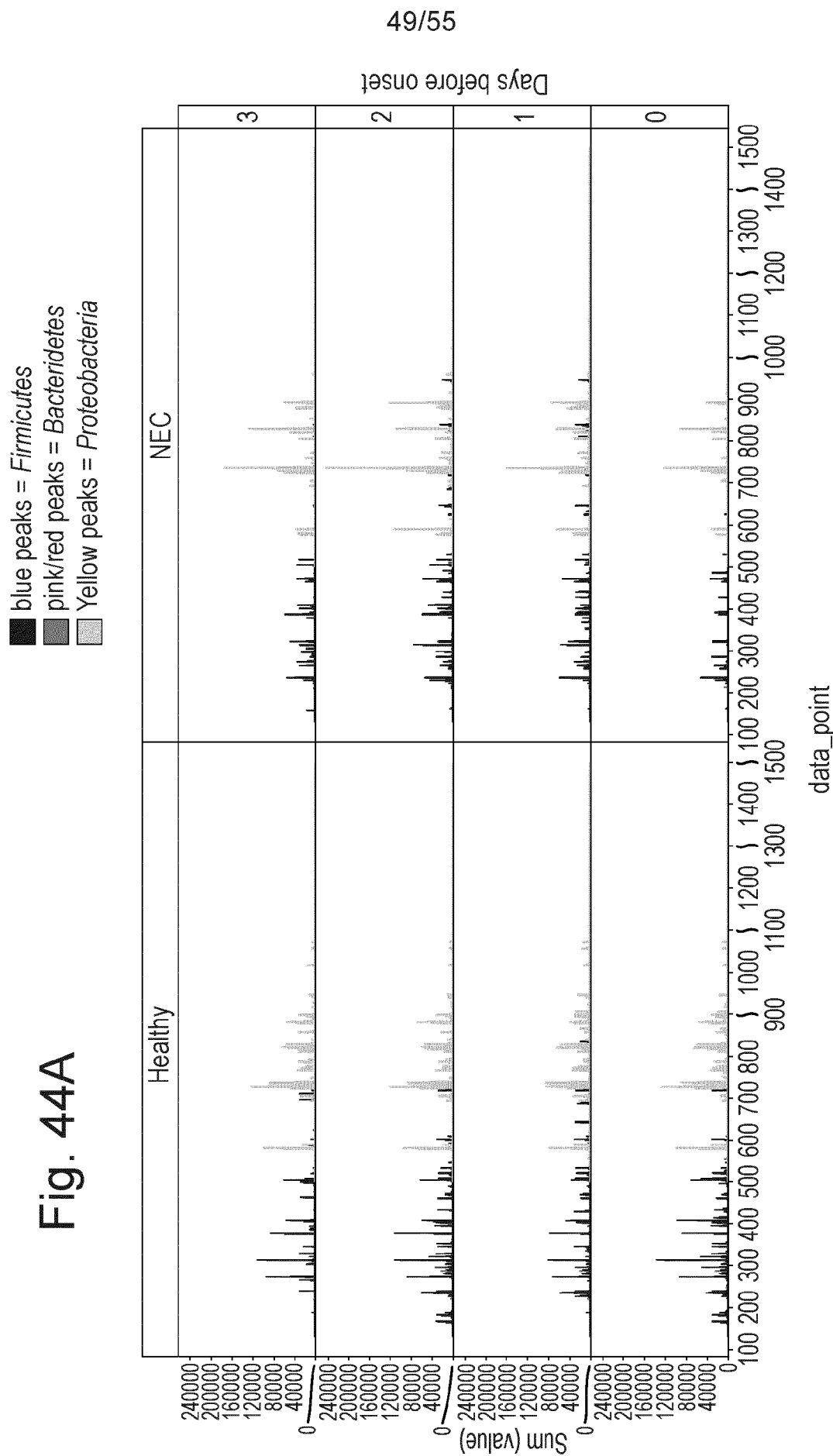
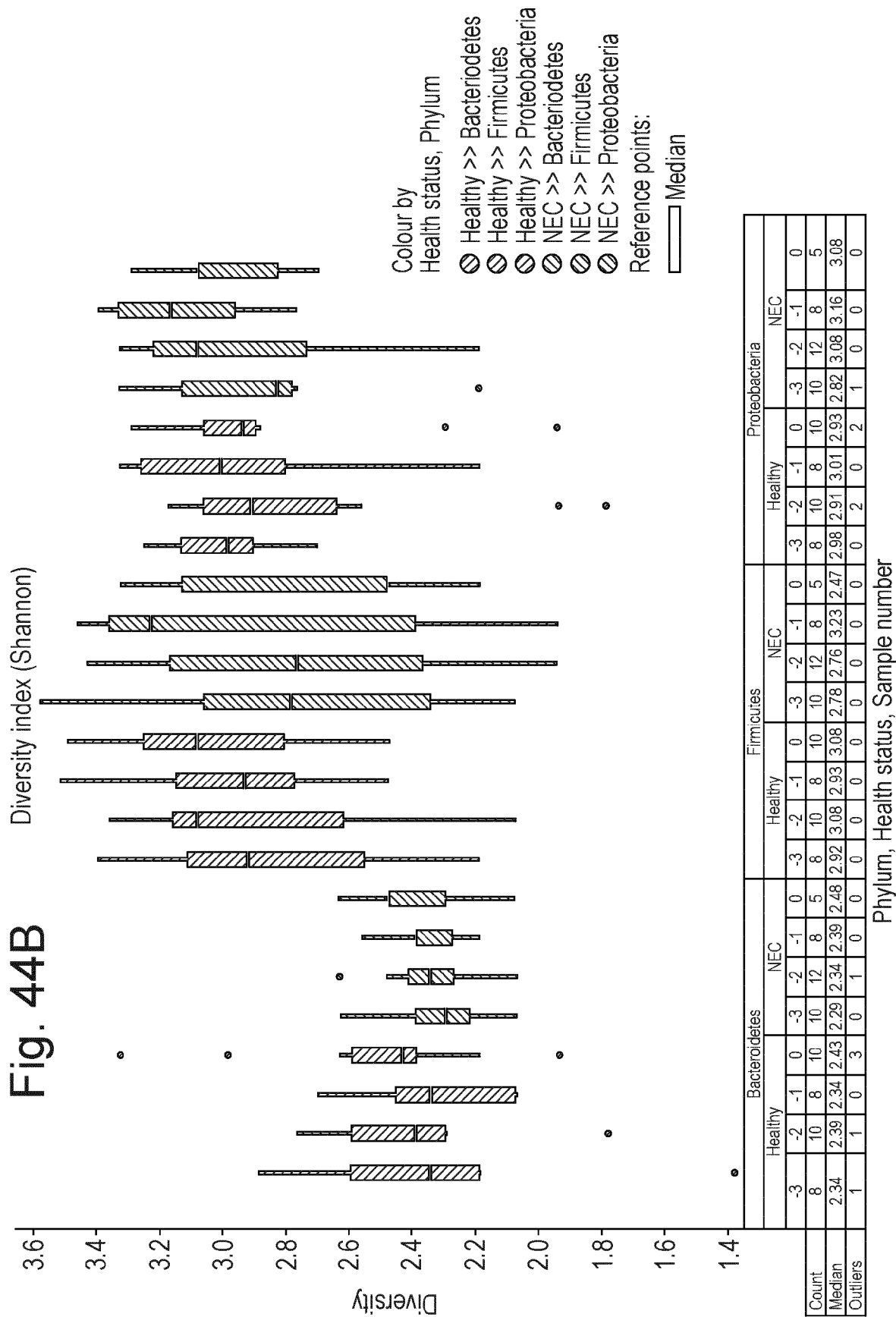
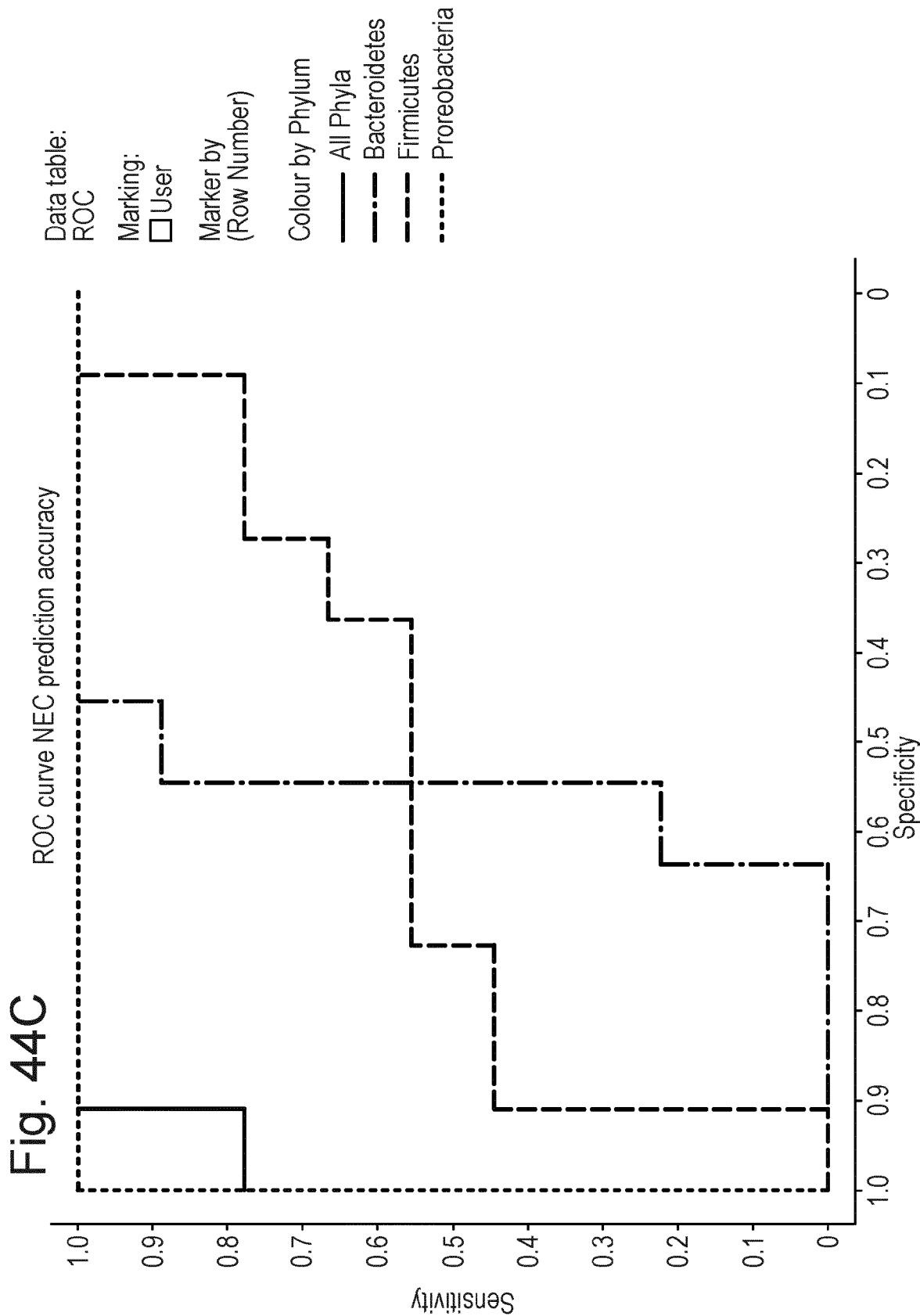
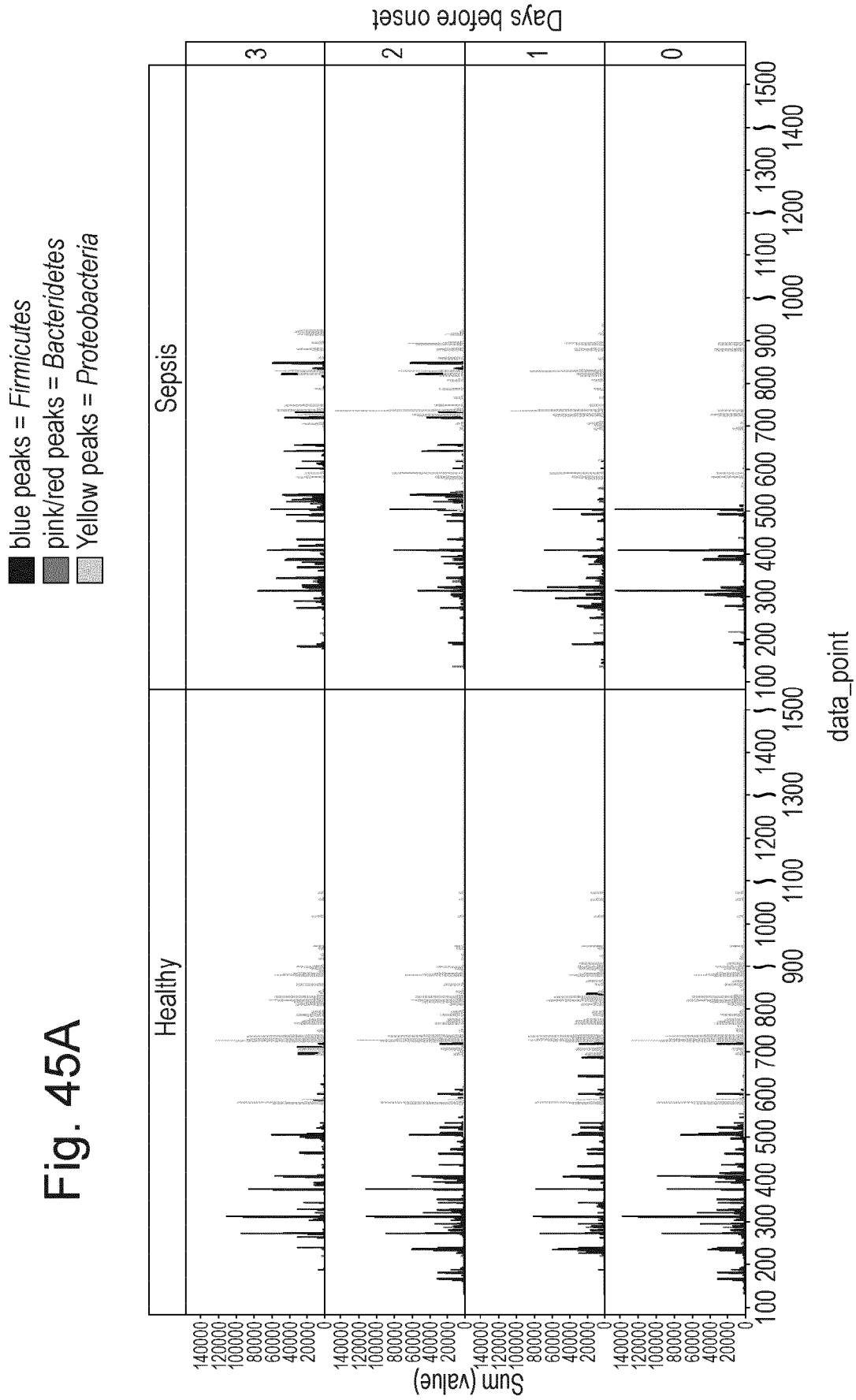


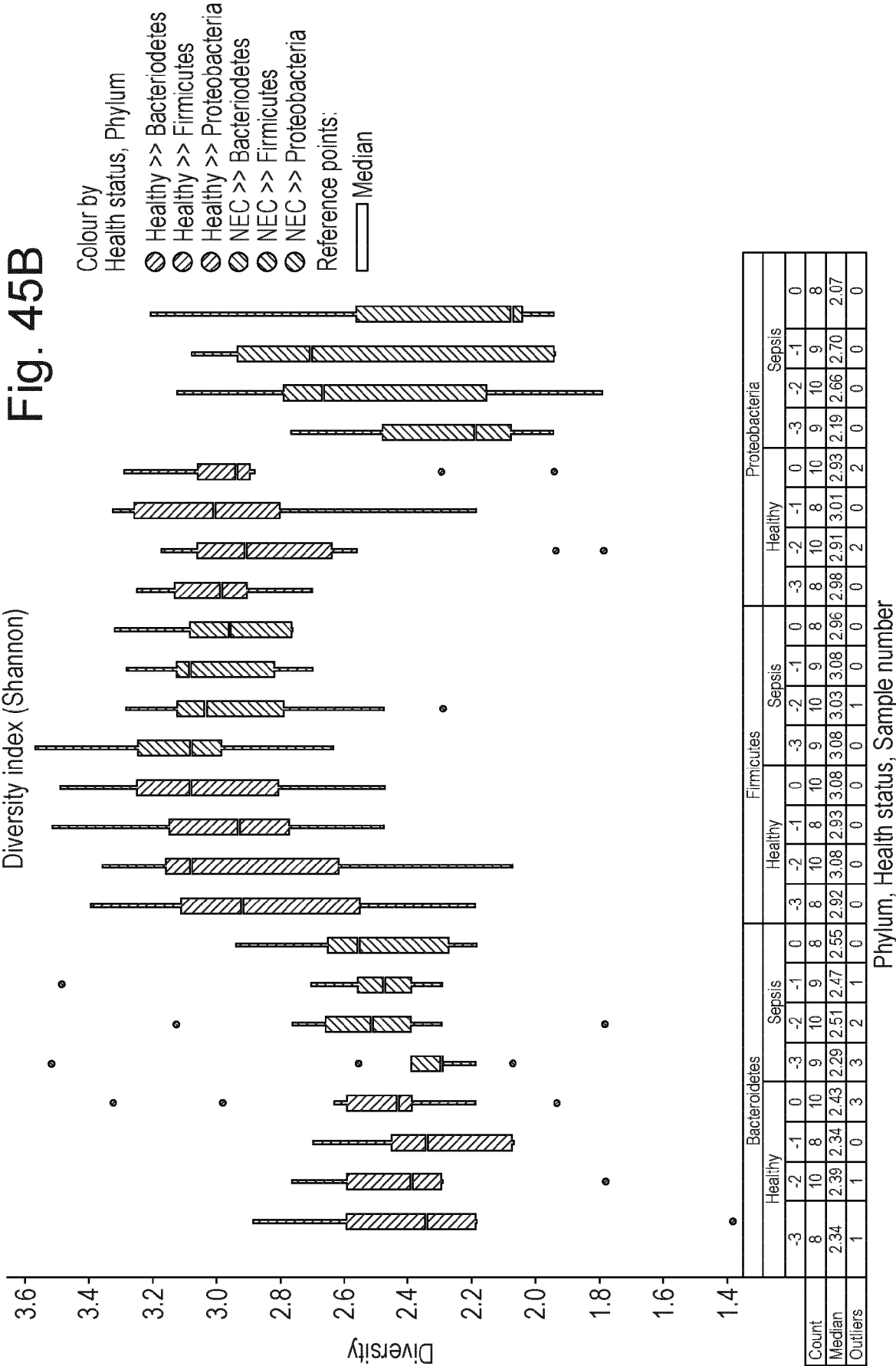
Fig. 44A





52/55





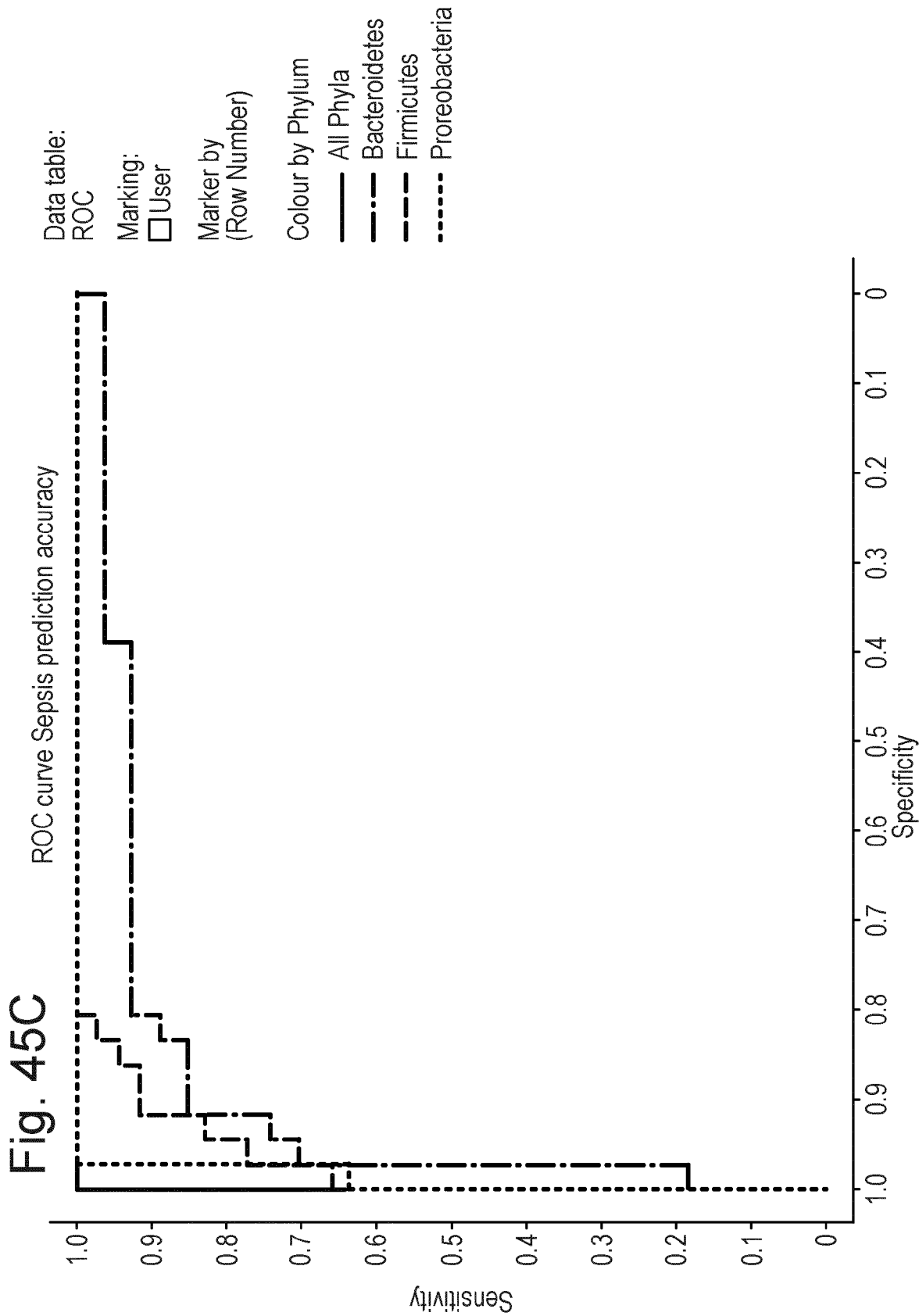


Fig. 46
Clostridium difficile associated diarrhoea
diagnosis with IS-pro

