



(51) International Patent Classification:

C12Q 1/68 (2006.01)

(21) International Application Number:

PCT/EP2014/077672

(22) International Filing Date:

15 December 2014 (15.12.2014)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

13199610.0 24 December 2013 (24.12.2013) EP

14150903.4 13 January 2014 (13.01.2014) EP

(71) Applicants: UNIVERSITE DE LIEGE [BE/BE]; Patent Department, avenue Pré-Aily, 4, B-3140 Angleur (BE).

QUALITY PARTNER S.A. [BE/BE]; rue Hayeneux, 62, B-4040 Herstal (BE).

(72) Inventors: DAUBE, Georges; Universite De Liege, Microbiologie des denrées alimentaires, Boulevard de Colonster 20 (Bât. B43BIS), B-4000 Liege (BE). TAMINIAU, Bernard; Universite De Liege, Microbiologie des denrées alimentaires, Boulevard de Colonster 20 (Bât. B43BIS), B-4000 Liege (BE). NEZER, Carine; Quality Partner s.a.,

Rue Hayeneux, 62, B-4040 Herstal (BE). DELHALLE, Laurent; QUALITY PARTNER s.a., Rue Hayeneux, 62, B-4040 Herstal (BE).

(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.

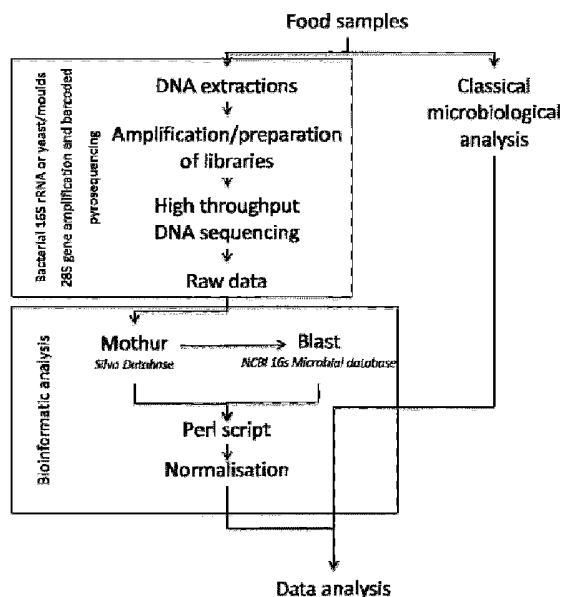
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, 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).

[Continued on next page]

(54) Title: METAGENOMIC ANALYSIS OF SAMPLES

(57) Abstract: A method of metagenomic analysis of a food or stool sample to detect and quantitate microorganisms present in said sample, the analysis comprising amplification of the V2- V3 region of bacterial 16S rRNA in said sample.

FIGURE 5





---

**Declarations under Rule 4.17:**

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

**Published:**

— *with international search report (Art. 21(3))*  
— *with sequence listing part of description (Rule 5.2(a))*

## Metagenomic Analysis of Samples

The invention relates to an analytical method that has been developed to identify and quantify the total microorganism content of any kind of samples including *inter alia* food samples and stool samples.

Microbial contents of food, faeces or environmental samples are frequently either identified or quantified through culture dependant methods. Among new generation molecular technologies the metagenomic analysis has emerged as a powerful tool. Metagenomics applies a suite of genomic technologies and bioinformatics tools to directly access the genetic content of entire communities of organisms. This technology is commonly used to describe microbial community profiles of various ecosystems like seas, soils, rumen, and faeces but also to study complex microbiota of foodstuff.

Several problems remain with metagenomic analysis of foodstuffs and other like samples. Firstly, it is necessary to provide a global metagenomic approach which is not dependant on the microbe's ability to grow on determined culture media at any define environmental conditions (e.g. pH, T°, Aw). Such an approach should provide for the description as well as quantification of the observed microorganisms, which is of critical needs in some environments, for example when food spoilage issues are to be explained.

Secondly, prior art methods of metagenomic analysis may be biased by the presence in the samples being tested of non-viable microorganisms. In many cases, particularly in the metagenomic analysis of food samples, it is important to be able to characterise and quantify only the live microorganisms in a sample. Prior art metagenomic techniques do not allow for such an analysis and results, when applied to food sample analysis may therefore be misleading.

Thirdly, prior art methods of metagenomic analysis may be biased by the presence of contaminating DNA present in the sample, which is amplified together with the target sequence of interest, but which does not provide useful information on the microbiota of the sample. For example chloroplast DNA or mitochondrial DNA may co-amplify with the target, where the target is the well known 16S rRNA gene.

Fourthly, prior art methods of metagenomic analysis may be biased by the presence of a large amount of DNA from bacteria or other microorganisms deliberately added to foodstuff as an ingredient or by the presence of a natural dominant species. Such DNA can mask the DNA of interest in a metagenomic analysis.

5

There is therefore a need for improved metagenomic analysis methods and tools. The invention overcomes these and other problems.

### Summary of the Invention

10

The invention therefore provides an integrated metagenomic approach developed to (I) identify and (II) quantify the micro-organism content of any samples, (III) distinguish if the observed DNA sequences derived from dead or alive germs, (IV) block the well known co-amplification of mitochondrial and plastid DNA with universal bacterial 16S primers, that might represent more than 90 % of the sequences in some food preparations, and (V) block the amplification of over-represented germs avoiding rare but important germs to be observed.

15

Our metagenomic analysis is a PCR-based approach targeting the highly conserved rDNA locus encoding for the 16S rDNA gene. This gene is commonly used for bacterial taxonomic classification as it contains a succession of nine hyper variable regions with well-conserved sequences in between. The conserved sequences are used to design universal bacterial primers flanking a polymorphic region chosen based on its high information content for a precise species identification of microorganism content.

20

In a first aspect, the invention provides a method of metagenomic analysis of a food or stool sample to detect and quantitate microorganisms present in said sample, the analysis comprising amplification of the V2-V3 region of bacterial 16S rRNA in said sample. In some embodiments the V1-V3 region of bacterial 16S rRNA is amplified. The inventors have identified that the V2-V3 region provides excellent hypervariability between common food spoilage flora, and therefore allows for bacteria likely to be present in food samples to be distinguished and quantitated.

25

In some embodiments, the use of V2-V3 of 16S rRNA, according to the invention, in a metagenomic analysis, allows for the identification and distinguishing of sequences from

30

35

bacteria of the families: *Acetobacteraceae*, *Aeromonadaceae*, *Bacteroidaceae*,  
*Burkholderiaceae*, *Carnobacteriaceae*, *Clostridiaceae*, *Comamonadaceae*,  
*Enterobacteriaceae*, *Enterococcaceae*, *Flavobacteriaceae*, *Lachnospiraceae*,  
*Lactobacillaceae*, *Leuconostocaceae*, *Listeriaceae*, *Moraxellaceae*, *Neisseriaceae*,  
5 *Pseudoalteromonadaceae*, *Pseudomonadaceae*, *Shewanellaceae*, *Streptococcaceae*,  
*Vibrionaceae* and *Xanthomonadaecea*.

In more detail, the invention allows for the detection and distinguishing of bacteria of the  
genera: *Acetobacteraceae* *Acetobacter*, *Acetobacteraceae* *Gluconacetobacter*,  
10 *Aeromonadaceae* *Aeromonas*, *Bacteroidaceae* *Bacteroides*, *Burkholderiaceae*  
*Burkholderia*, *Carnobacteriaceae* *Carnobacterium*, *Clostridiaceae* *Clostridium*,  
*Comamonadaceae* *Comamonas*, *Comamonadaceae* *Delftia*, *Enterobacteriaceae*  
*Escherichia*, *Enterobacteriaceae* *Pantoea*, *Enterobacteriaceae* *Serratia*, *Enterococcaceae*  
*Enterococcus*, *Chryseobacterium*, *Flavobacteriaceae* *Flavobacterium*, *Flavobacteriaceae*  
15 *Myroides*, *Lachnospiraceae* *Butyrivibrio*, *Lactobacillaceae* *Lactobacillus*,  
*Lactobacillaceae* *Pediococcus*, *Leuconostocaceae* *Leuconostoc*, *Leuconostocaceae*  
*Weissella*, *Listeriaceae* *Listeria*, *Moraxellaceae* *Acinetobacter*, *Moraxellaceae*  
*Psychrobacter*, *Neisseriaceae* *Alysiella*, *Pseudoalteromonadaceae* *Pseudoalteromonas*,  
*Pseudomonadaceae* *Pseudomonas*, *Shewanellaceae* *Shewanella*, *Streptococcaceae*  
20 *Lactococcus*, *Streptococcaceae* *Streptococcus*, *Vibrionaceae* *Aliivibrio*, *Vibrionaceae*  
*Photobacterium*, *Vibrionaceae* *Vibrio* and *Xanthomonadaecea* *Xanthomonas*.

Metagenomic analysis of the DNA present in a sample may provide a relative  
quantitative analysis of different microorganisms present. However in preferred  
25 embodiments the amount of some, or each of the microorganisms detected is normalised  
and expressed as a percentage of the colony forming units in the sample. This may be  
important in some embodiments since the absolute relative amounts of microorganisms in  
a sample may not be indicative of the spoilage potential of the sample or the shelf life  
limit. The invention therefore allows one to accurately qualify the quality of a food  
30 sample.

Normalisation of the quantity of microorganisms in a sample may be carried out by  
determining the percentage of microorganisms represented by at least one operational  
taxonomic unit (OTU) in the sample. This may be combined with determining the colony

forming units count (CFU) of the sample and normalising the percentage of organisms represented by said OTU to a proportion of the total viable CFU count of the sample.

Further quantitation of DNA in the sample may be provided, in some embodiments, by  
5 conducting taxa specific quantitative PCR (qPCR) on samples following metagenomic  
analysis. In such embodiments, the taxa chosen for the qPCR analysis are those taxa of  
interest identified by the metagenomic analysis. Any form of qPCR may be employed  
using any suitable gene, and any detection and quantification process. For example the  
16S rRNA gene may be amplified in qPCR tests. Different qPCR tests that might be  
10 employed include those involving an intercalating dye, and those involving a labelled  
reporter primer or probe. For example qPCR may be conducted using a hydrolysis probe  
(e.g. a Taqman probe), a molecular beacon probe, a pair of dual hybridisation probes, an  
amplifluor primer system, a scorpion primer system, a light-upon-extension system or a  
QZyme primer system.

15

In some embodiments the V2-V3 region of 16S rRNA is amplified in the metagenomic  
analysis as part of a larger fragment of the 16S rRNA gene. For example, in some  
embodiments, the V1-V3 region is amplified.

20 In some embodiments one or both primers having sequences comprising the sequences of  
SEQ ID NOs 1 and/or 2 may be used to amplify the V1-V3 region. Preferably, a pair of  
primers comprising the sequences of SEQ ID NO. 1 and SEQ ID NO. 2 respectively are  
used to amplify the V1-V3 region. In some embodiments the primers consist of the  
sequences of SEQ ID NOs 1 and 2 respectively.

25

Such primers may be modified by the addition of adaptors, key sequences, tags or  
universal sequences. In some embodiments, at least one of the primers has the sequence  
of SEQ ID NO. 82. In some embodiments, at least one of the primers as the sequence of  
SEQ ID NO. 83. In some embodiments a pair of primers having the sequences of SEQ ID  
30 NO.s 82 and 83 respectively is used.

In order to improve the analytical power of metagenomic analysis of a sample, the  
invention provides methods in which the amplification of plastid or mitochondrial DNA,  
and/or DNA from selected microorganisms is reduced. The invention therefore provides,  
35 in a further aspect, the use of Peptide Nucleic Acid (PNA) clamping in a method of

metagenomic analysis of a sample, to reduce amplification of at least one of: i) chloroplast DNA present in the sample; ii) mitochondrial DNA present in the sample; or iii) DNA from one or more selected microorganisms that may be present in the sample.

5 In some embodiments, the PNA clamping comprises conducting metagenomic analysis in the presence of a PNA oligonucleotide comprises a sequence derived from a chloroplast 16S rRNA gene, preferably a sequence common to a plurality of chloroplast 16S rRNA genes.

10 In some embodiments the PNA oligonucleotide comprises a sequence common to at least two of, and preferably all of, the chloroplast 16S rRNA sequences of SEQ ID NOs 3 to 10, or common to at least two of, and preferably all of, the chloroplast 16S rRNA sequences of SEQ ID NOs 18 to 25.

15 In some embodiments, the PNA oligonucleotide has a sequence not found in the bacterial 16S rRNA sequences of at least one of, and preferably all of, SEQ ID NOs 11 to 16, or has a sequence not found in the bacterial 16S rRNA sequences of at least one of, and preferably all of, SEQ ID NOs 26 to 31.

20 In some embodiments, the PNA oligonucleotide comprises the sequence of SEQ ID NO. 17. Preferably the PNA oligonucleotide consists of the sequence of SEQ ID NO. 17.

In some embodiments, the PNA oligonucleotide comprises the sequence of SEQ ID NO. 32. Preferably, the PNA oligonucleotide consists of the sequence of SEQ ID NO. 32.

25

In some embodiments the metagenomic analysis is conducted in the presence of a pair of PNA oligonucleotides comprising of or consisting of the sequences of SEQ ID NO.s 17 and 32 respectively.

30 In some embodiments, PNA clamping comprises conducting metagenomic analysis in the presence of a PNA oligonucleotide comprises a sequence derived from a mitochondrial 18S rRNA gene, preferably a sequence common to a plurality of mitochondrial 18S rRNA genes.

In some embodiments, the PNA oligonucleotide comprises a sequence common to at least two of, and preferably all of, the mitochondrial 18S rRNA sequences of SEQ ID NOs 33 to 38.

5 In some embodiments, PNA oligonucleotide has a sequence not found in the chloroplast 16S rRNA sequences, or the bacterial 16S rRNA sequences of at least one of, and preferably all of, SEQ ID NOs 39 to 47.

In some embodiments, the PNA oligonucleotide comprises the sequence of SEQ ID NO.  
10 48, optionally wherein said PNA oligonucleotide consists of the sequence of SEQ ID NO. 48.

In some embodiments, the PNA clamping comprises conducting metagenomic analysis in the presence of a PNA oligonucleotide comprises a sequence derived from a bacterial  
15 16S rRNA gene, preferably a sequence common to a plurality of bacterial 16S rRNA genes.

In some embodiments, the PNA oligonucleotide comprises a sequence common to at least two of, and preferably all of, the bacterial 16S rRNA sequences of SEQ ID NOs 49 to 61.  
20

In some embodiments, the PNA oligonucleotide has a sequence not found in the bacterial 16S rRNA sequences of at least one of, and preferably all of, SEQ ID NOs 62 to 79.

In some embodiments, the PNA oligonucleotide comprises the sequence of SEQ ID NO.  
25 80, optionally wherein said PNA oligonucleotide consists of the sequence of SEQ ID NO. 80.

In some embodiments the metagenomic analysis is conducted in the presence of any two, three or four of the PNA oligonucleotides discussed above, that is any two, three or four  
30 PNA oligonucleotides comprising or consisting of the sequences of SEQ ID NO.s 17, 32, 48 or 80 respectively.

The invention is further directed towards the ability to distinguish the amplification of DNA from live and dead microorganisms in a metagenomic analysis of a sample. The  
35 invention achieves this by the use of the intercalating dye propidium monoazide which



penetrates dead cells in the sample prior to DNA extraction, intercalates with the DNA of those dead cells and prevents subsequent amplification. The invention therefore provides, in a further aspect, a method of metagenomic analysis of a sample, wherein propidium monoazide (PMA) is added to the sample prior to DNA extraction in order to reduce the subsequent amplification of DNA from non-viable microorganisms present in the sample.

In each an every one of the uses of PNA clamping described herein or the use of PMA as described herein, the metagenomic analysis may be carried out in accordance with any other embodiment or aspect of the invention described herein.

10

In a yet further aspect, the invention provides a kit for metagenomic analysis of a sample comprising one or both primers comprising the sequence of SEQ ID NO. 1 and/or SEQ ID NO. 2, preferably wherein the primers consist of the sequences of SEQ ID NO. 1 and/or SEQ ID NO. 2.

15

In some embodiments, the primers of the kit may be modified by the addition of an adaptor, key sequence, tag or universal sequence. Preferably, the primers have the sequences of SEQ ID NO.s 82 and 83 respectively.

20

In some embodiments, a kit of the invention further comprises a PNA oligonucleotide comprising a sequence of any of SEQ ID NOs. 17, 32, 48 or 80, preferably wherein the primer consists of the sequence of any of SEQ ID NOs. 17, 32, 48 or 80. Preferably the kit may comprise a set of primers including any two, three or four PNA oligonucleotides comprising or consisting of the sequence of SEQ ID NO.s 17, 32, 48 or 80.

25

In some embodiments the kit of the invention may further comprise PMA.

In some embodiments the kit of the invention may further comprise a reagent for metagenomic analysis of a sample, e.g. a food or stool sample.

30

The invention will now be described in more detail with reference to the following illustrative Examples and the figures, in which:

Figure 1 shows the nine hypervariable regions of the 16S rRNA gene.

35

Figure 2 shows taxonomic information content along the whole 16S gene.

Figure 3 shows the sequence of *Escherichia coli* 16S ribosomal RNA gene (Genebank Accession Number AY804014.1, SEQ ID NO. 81) with annotation of the polymorphism frequencies across the bacterial kingdom.

Figure 4 shows primer statistics of both primers and amplicon product size, and the sequences of the forward and reverse primers (SEQ ID NO. 1 and SEQ ID NO. 2 respectively).

10

Figure 5 shows a flow diagram of the steps involved in the metagenomic analysis

Figure 6 shows an alignment of chloroplast 16S RNA sequences with various bacterial 16S rRNA sequences and the sequence of a first PNA clamp. Sequences entitled Chloroplast 1 to Chloroplast 8 are SEQ ID NO.s 3 to 10 respectively. SEQ ID NO.s of the other sequences are: *Pseudomonas* SEQ ID NO. 11; *Xanthomonas* SEQ ID NO. 12; *E. coli* SEQ ID NO. 13; *Lactococcus* SEQ ID NO. 14; *Brochothrix* SEQ ID NO. 15; *Arthrobacter* SEQ ID NO 16. the PNA oligonucleotide QP-PNAcup is SEQ ID NO. 17.

Figure 7 shows a further alignment of chloroplast 16S RNA sequences with various bacterial 16S rRNA sequences and the sequence of a second PNA clamp. Sequences entitled Chloroplast 1 to Chloroplast 8 are SEQ ID NO.s 18 to 25 respectively. SEQ ID NO.s of the other sequences are: *Pseudomonas* SEQ ID NO. 26; *Xanthomonas* SEQ ID NO. 27; *E. coli* SEQ ID NO. 28; *Lactococcus* SEQ ID NO. 29; *Brochothrix* SEQ ID NO. 30; *Arthrobacter* SEQ ID NO 31. the PNA oligonucleotide QP-PNAcdn is SEQ ID NO. 32.

Figure 8 shows an alignment of mitochondrial 18S RNA sequences with various bacterial 16S rRNA sequences and the sequence of a third PNA clamp. SEQ ID NOs. of the sequences shows are: *Mito-Oryza-sativa* SEQ ID NO. 33; *Mito-Oryza-rufipogon* SEQ ID NO. 34; *Mito-Bambusa-oldhamii* SEQ ID NO. 35; *Mito-Triticum-aestivum* SEQ ID NO. 36; *Mito-S.cereale* SEQ ID NO.37; *Mito-Sorghum-bicolor* SEQ ID NO. 38; Chloroplast E1 to Chloroplast E4 are SEQ ID NOs. 39 to 42 respectively; *E. coli* SEQ ID NO. 43; *Lactococcus lactis* SEQ ID NO. 44; *Weissella* SEQ ID NO. 45; *Brochotrix thermosphacta*

30

SEQ ID NO. 46; *Staphylococcus* SEQ ID NO.47. The PNA oligonucleotide QP-PNA<sub>m</sub> is SEQ ID NO. 48.

Figure 9 shows an alignment of 16S rRNA sequences from various bacteria of the lactococcus taxa, together with various contaminating bacteria, and a fourth PNA clamp. The SEQ ID NOs. of the sequences shown are: *Lactobacillus-delbrueckii-subsp* SEQ ID NO.49; *Lactobacillus-delbrueckii-subsp* SEQ ID NO. 50; *Lactobacillus-amylovorus* SEQ ID NO. 51; *Lactobacillus-acidophilus* SEQ ID NO. 52; *Lactobacillus-amyolyticus* SEQ ID NO. 53; *Lactobacillus-acetotolerans* SEQ ID NO. 54; *Lactococcus-piscium* SEQ ID NO. 55; *Lactococcus-plantarum* SEQ ID NO. 56; *Lactococcus-chungangensis* SEQ ID NO. 57; *Lactococcus-raffinolactis* SEQ ID NO. 58; *Lactococcus-lactis-subsp-Hordn* SEQ ID NO. 59; *Lactococcus-lactis-subsp-Lacti* SEQ ID NO. 60; *Lactococcus-lactis-subsp-Cremo* SEQ ID NO. 61; *Streptococcus-Thermophilus-19-5* SEQ ID NO. 62; *Streptococcus-Thermophilus-UKPS2* SEQ ID NO. 63; *Streptococcus-Thermophilus-SM1* SEQ ID NO. 64; *Streptococcus-parauberis* SEQ ID NO. 65; *Streptococcus-iniae* SEQ ID NO. 66; *Streptococcus-dysgalactiae-ssp* SEQ ID NO. 67; *Streptococcus-agalactiae* SEQ ID NO. 68; *Streptococcus-equi* SEQ ID NO. 69; *Streptococcus-oligofermentans* SEQ ID NO. 70; *Enterococcus-faecalis* SEQ ID NO. 71; *Lactobacillus-algidus* SEQ ID NO.72; *Lactobacillus-frumenti* SEQ ID NO. 73; *Lactobacillus-panis* SEQ ID NO.74; *Lactobacillus-fermentum* SEQ ID NO. 75; *Lactobacillus-oligofermentans* SEQ ID NO. 76; *Lactobacillus\_plantarum* SEQ ID NO. 77; *Brevibacterium-linens* SEQ ID NO. 78 *Psychroflexus-sp* SEQ ID NO. 79. The PNA oligonucleotide PNA<sub>Lactococcus</sub> is SEQ ID NO. 80.

Figure 10 shows metagenomic analysis of steak tartare samples.

Figure 11 shows relative proportions of the major bacterial taxa (over 1% of prevalence) in steak tartare samples.

Figure 12 shows a representation of the converted relative bacterial count of the OTUs in total count numbers of population. Only taxa whose population reached 10,000 cells per gram are visible. The bold bar indicates the threshold of 6 logs of CFUs above which the spoilage activity would be noticeable.

Figure 13 shows a comparison between bacterial content of 3 probiotics with and without PMA treatment.

Figure 14 shows an analysis of a pasta product with the standard metagenomic protocol,  
5 with the 2 PNA clamps specific to avoid plastid DNA, with the PNA directed against the mitochondrial DNA and finally with the 3 PNA clamps together.

Figure 15 shows bacterial profiles of 2 cheeses obtained with and without the use of a  
10 PNA Lactococcus clamping probe.

Figure 16 shows amplification of DNA from samples with and without PNA. In both the  
15 bacillus that are present in high concentration on the metagenomic profile are not amplified with the adjunction of the specific PNA design against a specific bacillus sequence.

Figure 17 shows amplification of DNA from samples with and without PNA. In both the  
20 bacillus that are present in high concentration on the metagenomic profile are not amplified with the adjunction of the specific PNA design against a specific bacillus sequence.

Figure 18 shows **Analysis A** shows the metagenomic result from a yoghurt product  
without any PNA use where more than 98 % of the sequences are coming from streptococcus thermophilus (94,9 %) and lactobacillus delbrueckii (3,6 %) leaving less place for the observation of real contamination floras.

25 **Analysis B** shows the metagenomic result with the use of the specific PNA blocking the amplification of any DNA originating from streptococcus thermophilus. In this case, the second technologic flora is taking the lead with the presence of 84,5 % of the sequences originating from lactobacillus delbrueckii.

30 **Analysis C** shows the metagenomic result with the use of the specific PNA blocking the amplification of any DNA originating from lactobacillus delbrueckii. . In this case, the second technologic flora is taking the lead with the presence of 98,6 % of the sequences originating from streptococcus thermophiles.

**Analysis D** shows the metagenomic result with the use of the specific PNA blocking the amplification of any DNA originating from both technologic species (streptococcus

thermophilus and lactobacillus delbrueckii). In this case only, the flora of interest is visible as a result of the metagenomic analysis of the food product.

5 Figure 19 shows a comparison of total flora estimated by classic microbiology and deduced by targeted qPCR for *Leuconostoc* sp, *Lactobacillus* sp and *Lactococcus* sp.

Figure 20 shows (A): Estimation (in percentage) of major taxon relative proportions by metagenomic analysis; (B): Estimation of each taxon concentration (in cfu/g) using the  
10 total flora estimation based on quantitative PCR.

## Examples

### Material and methods

15

#### *DNA isolation*

Twenty-five gram sample of meat product are homogenized for 1 minute in 225 ml in a tempo bag (bioMérieux Basingstoke, England, ref 80015) with sterile physiologic water  
20 using a stomacher apparatus. 1.5 ml of the suspension are then used for total genomic DNA extraction using a home made lysis step prior to the conventional use of the DNeasy Blood & Tissue Kit (QIAGEN, Crawley, UK). The DNA concentration is then evaluated by fluorometric dsDNA quantification using PicoGreen.

25 *Descriptive metagenomic analysis*

#### - Amplicon selection and primer designs

Culture-independent 16S rRNA gene sequencing has been widely applied to examine  
30 microbial diversity. Bacterial 16S ribosomal RNA (rRNA) genes contain nine "hypervariable regions" (V1-V9) (Figure 1), that demonstrate considerable sequence diversity among different bacteria. Unfortunately, 16S rRNA hypervariable regions exhibit different degrees of sequence diversity, and no single hypervariable region is able to distinguish among all bacteria. The full-length 16S rRNA genes (about 1500 bp) can  
35 be used for accurate taxonomic identification based on the underlying sequence diversity

among different bacterial species. However, the current high-throughput DNA sequencing technologies can only produce 16S rRNA gene fragments, instead of full-length genes. The challenge is to obtain reliable representation of bacterial communities through taxonomic classification of short 16S rRNA gene sequences

5

The inventors have discovered the most appropriate region for food metagenomic analysis. Information content of 16S rRNA, for bacterial assignment to the species level, were analyzed for different 16S regions across the most frequent food taxa.

10 For this purpose, reference sequences from the most common food bacterial taxa were extracted from open databases (Silva) and in silico cut in various fragment containing the sequences of a single or two contiguous hypervariable regions. After comparing those fragments to the databases, percentages of taxonomical identification to the species were computed for each region. As can be seen in figure 2, the V2-V3 region seems to be the  
15 more informative with respect to the common food spoilage flora.

Given the small number of base pairs contained between V1 and V2 and their low level of conservation, the forward primer was designed upstream of V1 (Figure 3).

20 The presence of polymorphisms in the selected primers might cause a failure to detect some bacterial species and consequently lead to an incomplete picture of the microbiote composition characterization.

For this purpose, different “in silico” analysis have been driven (data not shown) and  
25 have led to the degeneration of 2 bases in the 16S forward primer (figure 3).

The final oligonucleotide design included 454 Life Sciences A or B sequencing titanium adaptors (Roche Diagnostics, Vilvoorde, Belgium) and key sequence fused to the 5’ end of the specific target primers. The multiplex identifiers (MIDs) are only included along  
30 with the A adaptors.

As seen in Figure 3, two bases has been degenerated due to observed polymorphism across genera (results not shown)

35 16S gene library construction and pyrosequencing

The bacterial V1-V3 regions of the 16S rRNA gene were amplified by PCR using a homemade fusion primer set leading to an amplicon of 625 bp in length. For the GS Junior Titanium chemistry, the forward primer (5' - CCTATCCCCTGTGTGCCTTGGCAGTCTCAGGAGAGTTTGATYMTGGCTCAG -3') (SEQ ID NO. 82) contain the B adaptor, a key sequence (TCAG) and our forward conserved bacterial primer (Roche, Basel, Switzerland). The reverse primer (5' - CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNNNNNNTACCGCGGCTGCTG GCAC -3') (SEQ ID NO 83) contain the A adaptor, a key sequence (TCAG), ten bases barcodes (Ns) to distinguish samples and our reverse conserved bacterial primer.

The amplification mix contained 5 U of FastStart high fidelity polymerase (Roche Diagnostics, Vilvoorde, Belgium), 1x enzyme reaction buffer, 200 µM dNTPs (Eurogentec, Liège, Belgium), 0.2 µM of each fusionprimer and genomic DNA in a volume of 25 µl.

The PCR conditions are as follows: 4 min at 94 °C, 25 cycles of denaturation (15 sec at 94 °C), annealing (45 s at 56 °C) and extension (60 s at 72 °C), followed by a final elongation (8 min at 72 °C). These amplifications were performed on an Ep Master system gradient apparatus (Eppendorf, Hamburg, Germany).

The PCR products of 625 nucleotides are checked by gel electrophoresis and purified using the AMPure kit (Agencourt Bioscience Corporation, Beverly, USA) to remove amplicons shorter than 100 bp. Equal amounts of each of the PCR products are pooled and subsequently amplified by emulsion PCR before sequencing. Pyrosequencing is then performed with the Roche 454 GS Junior Sequencer (Roche, Basel, Switzerland) using the massively parallel pyrosequencing protocol.

#### Bioinformatic analysis

Image and data processing for amplicon sequencing are performed using the Genome Sequencer FLX System Software Package 2.3 (Roche, Basel, Switzerland). The Mothur program is used for the first part of the sequences analysis. Sequences with a length under 450 bases are discarded. The sequences with ambiguous bases or homopolymers longer than ten nucleotides and with an average quality score lower than 25 are removed

along with tag and primer sequences. Sequences from multiplexed samples are assigned based on the presence of the unique barcodes assigned to each sample. The sequences that pass the quality check are aligned to the SILVA alignment database. Chimeric sequences are detected using the chimera-slayer command included in the Mothur package, and the potential chimeras are removed. The sequences are preclustered to reduce false operational taxonomic units produced by erroneous sequences. A distance matrix is prepared, (distance= 0.01) and the sequences are clustered to operational taxonomic units (OTUs) using the average neighbour algorithm. An 80% confidence threshold is required for genus level assignment with the SILVA Database. OTU representing less than 0.01% of the total number of reads are considered as artifacts and discarded.

In a second phase, the representative sequences of each OTU are compared to the NCBI 16S microbial database for example using the Basic Local Alignment Search Tool (BLASTN). The BLAST assignments for each OTU are assigned if the BLASTN default score is above 800. The genus assignments obtained by Mothur and by BLAST are compared and noted as unclassified in a case of mismatch. In cases of several hits with a score above 800, the first hit with the highest score is fostered to give the bacterial species, and 5 following hits with a decreasing score are given for complementary information on the BLAST results.

Finally, conversion and normalization of data are carried out. The number of sequences from each OTU is used to compute percentages. The percentage of each OTU is finally converted as a proportion of the total viable count obtained by classical microbiological analysis. Figure 4 gives the flow diagram of the metagenomic analysis of any sample.

The complete description of the bioinformatic homemade pipeline is described in annexe 1 which is herein incorporated in its entirety.

### 30 ***Semi-quantitative metagenomic analysis***

Conversion of sequence proportions into real colony forming units is of great importance especially for the food sector. For a microorganism to induce organoleptic disorders, its average concentration has to reach minimum  $10^{5-6}$  CFU per gram of sample. Indeed, saying that a spoilage organism represents even 90 percent of the observed bacteria does



not mean anything if just a low amount of total flora is present in the product. Normalization of the sequences numbers in a log CFU estimation is therefore important.

Microbiological analyses are performed using a classic microbiological approach for the  
5 determination of the total bacterial count and, in some extent, the enumeration of bacterial sub-populations.

Ten-fold dilution series of the homogenized samples are prepared in peptone water and  
10 plated onto the corresponding medium. Total bacterial count is measured by counting the colony forming units (CFUs) present on plate count agar (PCA) (Oxoid, Dardilly, France) after 48 h of incubation at 30 °C. The results are expressed in logarithms of CFU per gram of sample.

The percentage of each OTU might then be converted as a proportion of the total viable  
15 count obtained by classical microbiological analysis.

As metagenomic analysis is a non-cultured based method, DNA derived from dead  
20 bacteria is also sequenced. Sequences normalization on the total growing bacterial count is introducing an underestimation. To counteract this bias, qPCR targeting diverse taxa may be provided.

### ***Discrimination between dead or alive bacteria***

Metagenomic analyses have revolutionized microbial detection. However the obstacle  
25 for DNA-based techniques is that DNA in the environment can be very stable and can persist for extended periods of time (days to weeks) after cell death. As DNA from dead or alive microorganisms is PCR amplified in the same way, this will lead to an overestimation of microbial targets. One of the prerequisites of making ecological conclusions derived from genetic fingerprints is that bacterial community profiles reflect  
30 the live portion of the sample of interest.

Sample treatment with Propidium monoazide (PMA), a DNA intercalating agent, prior to  
DNA extraction successfully avoids amplification of targets from dead microorganisms.

Propidium monoazide is a membrane-impermeant dye that selectively penetrates cells with compromised membranes, which can be considered dead. Once inside the cells, PMA intercalates into the DNA and can be covalently cross-linked to it after exposure to strong visible light. This subsequent complex strongly inhibits further PCR  
5 amplification. By using PCR after PMA treatment, the analysis of bacterial communities can theoretically be limited to cells with intact cell membranes.

Twenty-five gram of food sample are homogenized for 1 minute in 225 ml in a tempo bag (bioMérieux Basingstoke, England, ref 80015) with sterile physiologic water using a  
10 stomacher apparatus. 1.5 ml of the suspension are then centrifuge. After discarding the supernatant, the pellet is exposed for 10 minutes to PMA treatment (50  $\mu$ M) in a dark room.

After a 5 minutes light exposure, the solution is centrifuge again and the pellet is directly  
15 mixed with the lysis solution prior to DNA extraction using DNeasy Blood & Tissue Kit (QIAGEN, Crawley, UK).

***Elimination of plastid and mitochondrial DNA co-amplification.***

20 In case of prepared food samples containing spices, herbs, fungi, vegetables or fruits; whatever the extraction method used to isolate DNA from these their bacterial communities, the extracted microbial DNA is generally contaminated with DNA of plant origin, in particular chloroplasts and mitochondria. The same problem will occur for any 16S metagenomic analysis aiming to learn about plant-associates microbiotes.

25 Since the chloroplast 16S and mitochondrial 18S rRNA genes share high sequence similarity with bacterial 16S rRNA sequences contamination with plant DNA poses a serious challenge for the application of PCR-based methods to profile and quantify bacterial populations in plant environments.

30 Without specific treatment, it is not rare to sequence more than 95 percents of mitochondrial and chloroplastic DNA instead of the bacterial one that is actually targeted, which is quite irrelevant, and undesired considering the price of a metagenome analysis.

35

PNA (Peptide Nucleic Acid), an artificially created DNA analogue, was invented by Drs. Nielsen, Egholm, Berg, and Buchardt in 1991. The phosphate ribose ring of DNA was replaced with the polyamide backbone in PNA. Despite a radical structural change, PNA is capable of sequence-specific binding in a helix form to its complementary DNA or RNA sequence. Due to its superior binding affinity and chemical/biological stability, PNA has been widely applied in the field of biology ([www.panagen.com](http://www.panagen.com)).

PNA PCR clamping technique was developed very early to detect SNP in low abundance. For example, a few mutant cancerous cells may be among healthy cells in a tissue biopsy. PNA clamp complementary to wild type sequence hybridizes specifically with the wild type and blocks its PCR amplification while allowing amplification of mutant sequence of imperfect match. A single base mismatch is enough to discriminate amplification of mutant type from wild type.

Two PNA clamps have been developed to block the co-amplification of the chloroplastic DNA (Figures 6 and 7).

A third PNA clamp has been developed to block the co-amplification of the mitochondrial DNA (Figure 8).

The use of those 3 PNA clamps in the PCR mix allow to mainly amplify the desired bacterial 16S DNA for the microbiome profiling assay.

***Elimination of specific undesired over-represented taxa.***

In some cases, the bacterial diversity might be hidden by an over represented taxa. One example is the presence of a huge amount of *Lactococcus* species in any cheese as these taxa are extensively used as starter in the production of buttermilk and cheese. Specific PNA clamping probes have been designed to avoid the amplification of any bacteria belonging to the *Lactococcus* taxa in order to give emphasis on the environmental contaminating flora or on any other species contributing to the specific texture, flavour or odour of the cheese.

On the same principal of the mitochondrial or plastid design, we founded a PNA sequence between our metagenomic V1-V3 primers. The PNA probe is characterized by a

sequence conserved for example across the *Lactococcus* taxa but showing at least one point mutation compare to the other bacterial classes frequently observed in cheese products.

The alignment used for the lactococcus PNA probe is represented in Figure 9.

5

## Results

### *Descriptive metagenomic analysis*

10 The present invention might be used for a complete description of the total flora of any sample (food, beverage, stools, any biological sample, soil, water or any environment sample)

As an example of a descriptive metagenomic analysis, we analyse a set of steak tartare  
15 samples. Steak tartare is indeed a popular meat dish in Belgium and other European countries. This meat preparation, due to its raw nature, is highly sensitive to bacterial spoilage. A better understanding of the bacterial content of this product will thus be insightful to control the risk of spoilage. Metagenomics targeted on the 16S ribosomal DNA has appeared as a powerful tool to study bacterial composition of food samples.  
20 The aim of this study is to identify the bacterial populations of steak tartare from different origins along their shelf life and determine the main spoilage bacterial species.

Results of the metagenomic analysis on steak tartare are shown in Figure 10. Each lane represents 100% of the bacterial composition of one sample. Each colour corresponds to  
25 a certain amount of sequences assign to a certain taxa.

A total of 32 dominant bacterial species were identified in the steak tartare samples by the metagenomic analysis. *Brochothrix thermosphacta*, *Lactobacillus algidus*,  
*Lactococcus piscium*, *Leuconostoc gelidum*, *Photobacterium kishitanii*, *Pseudomonas sp.*  
30 and *Psychrobacter urativorans* are the most predominant bacterial species in the samples and represent 89% of the detected bacterial species.

*Brochothrix thermosphacta* and *Lactobacillus algidus*, represent 50% of all the detected species. *Brochothrix thermosphacta* is detected mostly in samples from restaurants (2,8-  
35 93,9 %; n=4/6) and supermarkets without intern butcheries (0,3-83,2 %; n=8/8).

*Lactobacillus algidus* is present in each category but in low proportion for restaurants (2,3-5,5 %, n=3/6). *Pseudomonas sp.*, *Brochotrix thermosphacta* or *Leuconostoc gelidum* are present in high proportion in some samples at Day 0. These bacteria are well known to spoil meat product and samples that contain these bacteria in high proportion are more susceptible to spoil rapidly. Some samples (data not shown) contain a high level of *Streptococcus* or *Enterobacter* resulting probably of cross contamination during the meat process.

A food with a concentration above 6 log CFU/g is generally considered as spoiled. However, the results indicate that some samples have a concentration around this theoretical limit and contains a high proportion of bacteria which are not considered as spoilage bacteria like *Lactobacillus algidus*.

The metagenomic analysis is a powerful tool to identify and to measure the relative proportions of dominant bacterial species in the steak tartare as any other food or non food samples.

Some bacterial species could be indicators of the meat quality. Identification of some bacterial species over time gives information on the evolution of the spoilage process.

Therefore, metagenomic analysis could be an additional tool to control and manage the meat quality as well as the risk of spoilage.

#### ***Semi-quantitative metagenomic analysis***

The main drawbacks of the metagenomic approach is the relativity of the proportions of the bacterial populations. The following example shows that linking these proportions to a validated estimated bacterial count is an actual tool to accurately qualify the quality of a food sample.

The figure 11 shows the relative proportions of bacteria found in several samples of steak tartare preparation obtained from 2 restaurants, 2 from sandwich vendors and 2 from butcheries. The metagenomic sequencing and analysis followed the same protocol as described above.

35

The samples are clearly populated with various populations with a leading domination of Lactic Acid Bacteria (LAB). The first taxon is *Lactobacillus algidus*, which is a common bacterium frequently found in cold stored beef meat. This bacterium is homofermentar and not known for any spoilage activity. Another LAB is *Lactococcus piscium* which again is a non-spoiling bacteria.

Other populations observed however are known to possess spoiling activity when their population is high enough. This is the case for *Leuconostoc*, sp., *Weissella* sp., *Photobacterium phosphoreum* and *Brochothrix thermosphacta*.

We would like to focus on 2 samples: Restaurant-2 and sandwich-2. These samples possess the highest levels of contamination by *Leuconostoc*, *Weissella* and *Photobacterium* in terms of relative proportions.

Based on these observations, one could argue that both samples are not of good quality because of the presence of spoilage bacteria (*Leuconostoc* giving a cheesy and slimy flavor on the meat). But, in fact these relative proportions don't reflect the true level of bacterial contamination.

In order to visualize the results from the point of view of a food microbiologist, we convert the OTU populations into estimated log CFUs, taking the following as an arbitrary statement: that the proportion of an OTU to the entire OTU population was related to the same proportion of log CFUs to the log of the total plate count (Figure 12). As we chose matrices actively invaded by bacteria, these were considered as active and thus correctly represented by the colonies grown on plates. This is not a canonical choice but it allows the selection of OTU populations with a size of over 100,000 cells; a threshold above which is said to lead to spoilage for most altering bacteria in foodstuff at the shelf life limit.

In the sample restaurant-2, several spoilage bacteria are present with a population level above 1,000,000 cells per gram. Such level clearly has a noticeable sensory effect. On the contrary, the sample Sandwich-2 is in fact lowly contaminated. As the total flora count of mesophilic bacteria is below 10,000 cells per gram, it doesn't appear on the graph. Thus the spoilage population, event if they are present, will not have a sensory effect at the time of sampling.

This combination allows us to clearly see the true impact of the bacteria present.

***Discrimination between dead or alive bacteria***

5

Metagenomics is an extremely valuable tool for probiotics quality control. The methodology of the invention enables the validation of the determined bacterial and/or yeast content of the product. As described above, by adding a PMA treatment to the sample prior to the DNA extraction, the viability of the determined micro-organisms might be assessed.

10

As seen in figure 13, the proportions of some bacterial groups are modified by the PMA treatment. By comparison between a sample PMA treated ( $Px^{PMA}$ ) or not ( $Px$ ), one can see that the total amount of certain species is decreasing after a PMA treatment meaning that the analysis conclude positively about the presence of such a bacterial group in the probiotic but underline the fact that the viability of this group is not well conserved. Indeed, the decrease of proportion indicate that part of those bacteria are actually already dead which is not interesting as the effect of a probiotic intake will be due to the gainful activity of the viable bacteria.

20

***Elimination of plastid and mitochondrial DNA co-amplification.***

Since the chloroplast 16S and mitochondrial 18S rRNA genes share high sequence similarity with bacterial 16S rRNA sequences, contamination with plant DNA poses a serious challenge for the application of PCR-based methods to profile and quantify bacterial populations in plant environments. Microbiome profiling of a food product mixed with herbs, spices or any vegetable is really challenging without specific treatment avoiding the amplification of chloroplastic or/of mitochondrial DNA. Without this kind of treatment, the proportion of those “contaminating” DNA could indeed represent in some cases until 99 percents of the reads obtained by the high-throughput sequencing. Considering the cost of an analysis, sequencing 9900 sequences out of 10000 coming out of spices when trying to analyze the bacterial diversity of the product is definitely not what is expected.

30

We analyze such a product at the date of production normally meaning with a low amount of bacterial flora. Without any treatment, the majority of the DNA sequences were coming from chloroplastic DNA. Based on the bacterial and chloroplastic DNA alignment described earlier, 2 PNA clamping probes were design to block the amplification of any chloroplastic DNA during the 16S specific PCR. Using those probes, no DNA was amplified from chloroplast but the majority of the produced sequences were from mitochondrial DNA. On the same bases, we then design a PNA clamping probe directed against the amplification of any mitochondrial DNA. By combining the 3 PNA clamping probes, we could finally observe the bacterial diversity of the product (figure 14)

Without those probes, the bacterial diversity of many food or agricultural samples are just impossible to explore.

#### 15 *Elimination of specific over-represented taxa.*

In fermented food, the use of a starter to induce the fermentation has often as a result that those microorganisms that are part of the recipe characterize a major part of the final ecosystem of the product.

20

In case of quality problem leading to a non-conform product, it is really important to be able to identify the microbial diversity often hidden by the major starter microorganisms. As for vegetable products, the invention is using a PNA clamping probe to block the amplification of those “contaminant” flora to be able to see the diversity behind.

25

Figure 15 represents the results of a metagenomic analysis of two different cheese cords with and without PNA probe designed against the amplification the lactococcus taxa representing the major proportion of their microbial flora.

30 *Lactococcus lactis* is introduced thought the starter to initiate the cheese fermentation.

In both cheeses, the bacterial profile is dominated by the *lactococcus* genera preventing the underlying diversity to be observed. As can be clearly seen in figure 15, the use of the PNA clamp designed to block the amplification of any *Lactococcus* species has a huge effect on the observed bacterial profile. When PNA *Lactococcus* is used in the



metagenomic protocols, the *Lactococcus* genus clearly disappeared in favour of the underlying diversity.

## 5 Conclusion

Metagenomic analysis offers a new tool for identifying microorganisms present in perishable food and their evolution depending of environmental conditions. The information available through metagenomics analysis provides a clear picture of the microbial community. Microbiological ecology studies have shown that the bacterial world of foods are much more diverse than the cultivable group of bacteria studied by culture media. For the metagenomic analysis, 3 to 5 thousand sequences appeared largely sufficient to identify and to characterize the spoiling microflora, though that might change depending on the objective and the budget of a particular study. Many food manufacturers, government agencies, retailers, distribution quality laboratories and researchers use general culture media without precisely knowing the bacterial communities inside the food. Compared to culture based methods on selective media and previous independent culture techniques, metagenomic analysis combined with the enumeration of psychrotrophic flora gives more valuable information, and its use should be considered as a technique for quality control, for accurately determining shelf life or for developing new food products.

The invention relates to a global analytical tool that is especially useful in the food industry to identify and quantify the spoilage microbiota responsible of organoleptic alterations of the food.

The term metagenomics could be defined as "the application of modern genomics techniques to the study of communities of microbial organisms directly in their natural environments, bypassing the need for isolation and lab cultivation of individual species" Conventional sequencing begins with a culture of identical cells as a source of DNA. However, early studies revealed that there are probably large groups of microorganisms in many environments that cannot be cultured and thus cannot be sequenced.

For environmental studies, the interest in metagenomics is to analyse the complete bacterial diversity to determine the type and relative abundance of any populations present in heterogeneous samples, such as soil, marine, or gut microbiome.

The present invention may use metagenomic analysis as a tool for food safety and/or for food quality control, The present invention relates to a package of different analytical options allowing analysis to only focus on the viable spoilage flora specific to a particular kind of food sample. The present methods can be used whether the sample is:  
5 raw or cooked; with or without addition of spices or any kind of vegetables in the recipe; with or without the use of a bacterial starter to boost the fermentation process; and/or with or without added protective flora for extension of the product shelf life.

Unlike environmental samples, food samples are potentially subjected to different treatments like: Thermic treatment such as cooking, sterilisation, pasteurisation.

10 Those thermic treatments will kill most of the bacterial flora present in the food but the DNA will still be present and taken into account in the metagenomic description of the potential alteration cause of the cooked product and could induce an false interpretation of the spoilage origin.

Those cooked products should be more or less free of spoilage bacteria but could be  
15 subject to a post thermic treatment contamination.

#### Fermentation

A starter culture can be defined as a microbial preparation of large numbers of cells of at least one microorganism to be added to a raw material to produce a fermented food by accelerating and steering its fermentation process.

20 The use of starter culture shortens the fermentation process and reduces the risk of fermentation failure.

As a consequence, if the producer requires a metagenomic analysis of fermented food (cheese, kefir, yogurt, salami, wine, beer, sourdough,...) to check for contamination or spoilage flora, the present methods are able to detect sequences of the flora of interest  
25 through the use of PNA specifically added in the PCR reaction to block the amplification of unwanted DNA originating from the starter flora.

Addition of flora for a biological preservation of the food.

The addition of a large amount of protecting flora, that is bacteria not associated to any spoilage process in a specific kind of food sample, will contribute to decrease the  
30 development of the unintended spoilage flora.

Our invention will in this case also give the opportunity to cancel the amplification of the DNA originating from the protective flora through the use of specifically targeted PNA.

Addition of spices or vegetables in the recipe

In prepared food with added spices or vegetables in the recipe, the sequences of spoilage bacterial 16S rDNA are often hidden under an overabundance of sequences produced by co-amplification of the chloroplastidial and mitochondrial 16S rDNA.

5 The presence of mitochondrial and plastid DNA in those kinds of samples has limited the full realization of the potential offered by the last generation of metagenomics approaches dedicated to the feed and food industry.

The present invention will in this case also give the opportunity to cancel the co-amplification of the DNA originating from spices and/or vegetables through the use of specifically targeted PNA.

10 The present invention provides for the quantification of the identified spoilage bacteria by the use of our metagenomic analysis.

The food product contains some bacteria that are involved in spoilage, the so-called specific spoilage organisms (SSO), and others which grow without causing unpleasant changes. Depending of the food matrix, it is usually assumed that the spoilage process  
15 begins with a concentration of total flora higher than 5-6 log cfu/g. The SSO may be present at a low concentration and, in that case, are not able to spoiled the product.

The result of a metagenomic study is the identification of the different observed bacteria with their relative proportions and is therefore only semi-quantitative.

To correctly interpret the data from the ultra-sequencing it is crucial to develop a parallel  
20 method to estimate the concentration of the different bacterial species present in the food sample.

If coupled with the results of enumeration of total bacteria in conventional culture medium, the relative proportions obtained by metagenomic analysis might give an estimation of the real proportion of the differents observed species in terms on cfu per  
25 gram of samples.

However, one drawback of conventional microbiological methods lies in their failure to detect some of the existing microflora present in a food matrix resulting in an underestimation of the contamination levels of microorganisms.

30 To get an accurate and reliable counting of the metagenomic profile, the presetn invention relates to the use of newly developed quantitative real-time PCR (qPCR) of the various taxa predominantly present in food samples.

The present invention relates therefore to the estimation of the real amount of the different taxo identified in the metanenomic profile based of the real quantification of

one of them by qPCR giving an estimate of bacterial concentration by far more accurate and reliable..

### PNA

5 A food technological flora is a flora that brings more value to the product (texture, flavoring, acidification, nutritional properties, stabilization and conservation). This flora is often added during the manufacturing process of the food mastered in proportions so as to achieve the desired result in the finished product. In the food industry adding useful flora is a very common process.

10 This use might be a problem when the spoilage origin is to be analyzed through the use of a food metagenomic analysis. Indeed, if a food has been altered, the purpose of the metagenomic analysis will be to provide a description of the flora that might be responsible of this alteration. In the case of any use of technological flora during the process, the DNA of it might be present at more than 95 % preventing the spoilage flora' DNA from being observed.

15 In a food metagenomic analysis, the use of specific PNA acting against the amplification of any "contaminant" DNA belonging to any specific technologic flora is therefore importance.

### 20 PNA Bacillus

In spite of spore forming strains, the genus bacillus is widely used as starter in food fermentation. In Africa, *Bacillus subtilus*, *Bacillus pumilus* and *Bacillus Licheniformis* are involved in bean fermentaion. In Japan, Natto, a unique traditional health food, is made using a fermentation process by adding beneficial bacteria, *Bacillus natto* to soybeans.

25 The present invention relates to the development of a PNA blocking the amplification of any bacillus species used as or within a technologic flora.

The accuracy of the PNA design has been checked with the PNA designer of "Life Technologies". (<http://www6.appliedbiosystems.com/support/pnadesigner.cfm>)

30

35

Table 1

## Calculation Results

Sequence	Lys-AGA CTG GGA TAA CTC CG
N Terminus Reporter	N: None
C Terminus Reporter	C: None
Tm	[1 $\mu$ M] = 80 °C [1 nM] = 69 °C [1 pM] = 59 °C
Molecular Weight	4782,27
Lenght	Residues: 18 Bases: 17
% Purine	58,80%
Purine Strech	4
Complementarity	Moderate AGACTGGGATAACTCCG
Design Recommendations	Design Suitable

5 **PNA Streptococcus thermophilus + PNA Lactococcus delbrueckii**

In the milk processing industry, starter cultures and ferments are widely used particularly in the manufacture of yoghurt to provide a particular organoleptic aspect (acidification, flavour, texture etc...)

10 According to the Codex Alimentarius (A-11a of 1975), yogurt is a coagulated milk product resulting from the fermentation of fresh milk or pasteurized milk, with or without other added ingredients, thanks to the action of *Lactobacillus delbrueckii* subsp. *Bulgaricus* and *Streptococcus thermophilus*. In the finished product to have the “yogurt” designation, those 2 bacteria must be viable and abundant in quantity.

15 However, microbial diversity of such products may be masked by the presence of a large proportion of this technological flora. Thereby investigation of the diversity through the use of molecular techniques may be hindered by the very high concentration of these two

technological flora. It is therefore necessary to find or develop an adequate technique to circumvent this problem. The easiest way to address this problem is the exclusion at the analytical level of technological flora so that other flora, present at lower levels can be detected and to provide a more complete mapping of the microbial alteration of the product. The negative aspect of this approach is that the majority (sometimes > 99 %) of the generated sequences will come from those 2 bacteria and will therefore be irrelevant to explore the alteration of the product.

An other way is to block the specific DNA amplification of those 2 bacteria by adding 2 PNA molecules in the PCR reaction.

#### 10 Design PNA Streptococcus thermophiles

The accuracy of the PNA design has been checked with the PNA designer of “Life Technologies”. (<http://www6.appliedbiosystems.com/support/pnadesigner.cfm>).

Table 2

15

#### Calculation Results

Sequence	Cys-AAC CTG CCT TGT AGC GGG
N Terminus Reporter	N: None
C Terminus Reporter	C: None
Tm	[1 $\mu$ M] = 84 °C [1 nM] = 74 °C [1 pM] = 64 °C
Molecular Weight	5015,5
Lenght	Residues: 19 Bases: 18
% Purine	50,00%
Purine Strech	3
Complementarity	Low AACCTGCCTTGTAGCGGG
Design Recommendations	Design Suitable

Design PNA Lactobacillus delbrueckii

The accuracy of the PNA design has been checked with the PNA designer of “Life  
5 Technologies”. (<http://www6.appliedbiosystems.com/support/pnadesigner.cfm>).

Table 3

**Calculation Results**

Sequence	Lys-TAA CAA CAT GAA TCG CAT
N Terminus Reporter	N: None
C Terminus Reporter	C: None
Tm	[1 $\mu$ M] = 74 °C [1 nM] = 65 °C [1 pM] = 56 °C
Molecular Weight	5000,57
Lenght	Residues: 19 Bases: 18
% Purine	55,60%
Purine Strech	3
Complementarity	Moderate TAACAACATGAATCGCAT
Design Recommendations	Design Suitable

10

**The use of quantitative PCR to deduce total flora and metagenomic data.**

In prior art methods the relative proportions obtained by metagenomic analysis have been  
15 coupled with total flora enumeration from classic microbiology. This allows an estimation of the quantity of the major microorganisms present in a food sample.

However, these methods cause some biased results because numerous microorganisms cannot be cultivated. Enumeration data are thus widely misestimated.

To get a precise and reliable enumeration, it's important to work out an alternative technique to classical microbiology by quantitative PCR to detect and quantify some major taxons from foodstuffs.

The list of taxons to identify by qPCR has been obtained from previous metagenomic analyses of data on various food samples. These taxons were chosen based on their appearance frequency and their impact on the organoleptic qualities of the food.

The 7 taxons mainly present in classic perishable foodstuffs selected are:

- *Lactobacillus* sp.
- *Lactococcus* sp.
- *Leuconostoc* sp.
- *Psychrobacter* sp.
- *Pseudomonas* sp.
- *Brochrohtrix* sp.
- *Photobacterium* sp.

All the target genes used for the 7 qPCR tests belong to the conserved "housekeeping" genes, present in single copy in the bacterial genome, an important feature for the quantification method, allowing the design of "genus" common primers and hydrolysis probes allowing the relative proportion estimate of a majority of species included in a same bacterial genus (table 1).

Table 4 : Listing of primers and probes designed for the qPCR tests allowing the relative proportion estimate of 7 taxons mainly present in classic perishable foodstuffs :

target bacterial genus	target gene	primers	sequence
Lactobacillus sp	tuf	Lactobacillus-Tuf-F2	5'-GCYCACGTWGAATAYGAAAC-3'
		Lactobacillus-Tuf-R2	5'-CGDACTCCATTTCAACYAAGTC-3'
		Lactobacillus-Tuf-FAM1	5'-TGTGGCATWGGRCCATCAGTTGC-3'
Lactococcus sp	rpoA	Lactococcus-RecA-F2	5'-GCCGAAATYGATGGYGAAAT-3'
		Lactococcus-RecA-R2	5'-CAACTTTTTCACGCAATTGGTTG-3'
		Lactococcus-RecA-FAM4	5'-TGATGTCWCAAGCYATGCGTAAAC-3'



Leuconostoc sp	fus	Leuconostoc-Fus-F1	5'-TTCTTGTTCCATGAAATCCATTG-3'
		Leuconostoc-Fus-R1	5'-GAATACCCACTAGAWCGTACAC-3'
		Leuconostoc-Fus-FAM1	5'-TGTGTTTCACCAATTTGTGAATTTTACC-3'
Photobacterium sp	rpoA	photobacterium-rpoA-F1	5'-AATGCCAGGTTGTGCTG-3'
		photobacterium-rpoA-R1	5'-GACTTGTTTCAGAGTWAGAATWACTTC-3'
		photobacterium-rpoA-FAM1	5'-ATCGAYGGTGTGTTACATGAGTACA-3'
Pseudomonas sp	rpoA	Pseudo-rpo-F2	5'-GGCGTKATCGAYTCSGT-3'
		Pseudo-rpo-R2	5'-CAGMGGRCGCTGGTTGATG-3'
		Pseudo-rpo-FAM2	5'-TCGYGTTGCYGAYGACGA-3'
Carnobacterium sp	rpoA	Carnobacterium-rpoA-F1	5'-ATTGGYGTATTACCAGTCGA-3'
		Carnobacterium-rpoA-R1	5'-AACCATCTGCCCATACATC-3'
		Carnobacterium-rpoA-FAM1	5'-CGATTACACCCCAGTTAGTCGT-3'
Brochothrix sp	tuf	Brochothrix-tuf-F1	5'-CAGTACCTTCTGGTAACGTG-3'
		Brochothrix-tuf-R1	5'-ACTTCTTCACTGCAGAAGTATAC-3'
		Brochothrix-tuf-FAM	5'-TATTATCTAAAGAAGAGGGTGGACGTCATA-3'

### **Conventional PCR protocol**

Real-time PCR was performed using the LightCycler 480 system (Roche, Basel, Switzerland). The PCR reaction mixtures were placed in a 12 µl final volume containing 6 µl of LC480 probe master mix (Roche, Basel, Switzerland), 2 µl of template DNA (at 5ng/µl), 0,25 µl of primer pair (10µM each), 0,125 µl of Taqman probe (10µM). The reaction conditions included the initiation step for 10 min at 95°C, followed by 40 cycles of 15s at 95°C and 1 min at 60°C. The real-time system is supplied with the Lightcycler 480 Software version 1.5 using unique Roche algorithms for highly accurate and robust automated data analysis. Serial dilutions (10<sup>6</sup> to 1 copy numbers) of standard DNA are used for determining bacterial number. The mean of Cycle Threshold (CT) on the three repetition was used in order to estimate the load of targets bacterial populations present in the sample.

### **Specificity and sensitivity test.**

The specificity of each genus qPCR were tested on close bacterial species which may be present in foodstuffs to verify if each test does not interact with other bacterial species phylogenetically close (table 5).

In sensitivity test of the specific primers and probes, PCR amplifications were examined between specific primers and 10-fold serial dilutions of DNAs isolated from pure culture of 3 target species belongs to the target bacterial genus ranging from 10<sup>6</sup> to 10 copy numbers. The standard curves for all the 6 detection systems were constructed by using the Cp value obtained from the corresponding DNA concentration.

Table 5: 7 genus qPCR validation (A) sensitivity test ; (B) specificity test (ND= not detected).

10 (A)

	DNA copy numbers tested	Average Cp value ±SD for <i>Brochothrix</i> sp		
		<i>B.thermophacta</i>		
10-fold serial dilutions of DNAs	1,00E+06	19,32 ± 0,04		
	1,00E+05	23,46 ± 0,06		
	1,00E+04	27,04 ± 0,29		
	1,00E+03	31,09 ± 0,09		
	1,00E+02	35,27 ± 0,55		
	1,00E+01	39 ± 0,9		
	1,00E+00	ND		
linear regression parameters	efficiency	2,014		
	slope	-3,29		
	Y intercept	38,78		
	DNA copy numbers tested	Average Cp value ±SD for <i>Carnobacterium</i> sp		
		<i>C.maltomaticum</i>		
10-fold serial dilutions of DNAs	1,00E+06	18,05 ± 0,011		
	1,00E+05	21,67 ± 0,08		
	1,00E+04	25,27 ± 0,04		
	1,00E+03	28,69 ± 0,03		
	1,00E+02	31,8 ± 0,1		
	1,00E+01	34,28 ± 0,68		
	1,00E+00	38 (1/3)		
linear regression parameters	efficiency	1,907		
	slope	-3,56		
	Y intercept	39,47		
	DNA copy numbers tested	Average Cp value ±SD for <i>Leuconostoc</i> sp		
		<i>L.citream</i>	<i>L.gelidum</i>	<i>L.mesenteroides</i>
10-fold serial dilutions of DNAs	1,00E+06	18,52 ± 0,05	17,52 ± 0,02	16,89 ± 0,13
	1,00E+05	21,84 ± 0,05	21,08 ± 0,12	20,7 ± 0,02
	1,00E+04	25,56 ± 0,17	24,33 ± 0,13	23,99 ± 0,04
	1,00E+03	28,52 ± 0,08	27,67 ± 0,13	27,44 ± 0,13
	1,00E+02	31,06 ± 0,09	30,51 ± 0,18	30,33 ± 0,26
	1,00E+01	33,7 ± 0,14	32,55 ± 0,19	31,8 ± 0,29
	1,00E+00	ND	ND	ND
linear regression parameters	efficiency	1,997	1,997	1,944
	slope	-3,33	-3,382	-3,464
	Y intercept	38,56	37,85	37,83
	DNA copy numbers tested	Average Cp value ±SD for <i>Carnobacterium</i> sp		
		<i>Pangustum</i>		
10-fold serial dilutions of DNAs	1,00E+06	19,5 ± 0,22		
	1,00E+05	22,50 ± 0,23		
	1,00E+04	25,53 ± 0,13		
	1,00E+03	29,41 ± 0,07		
	1,00E+02	32,57 ± 0,11		
	1,00E+01	36,26 ± 0,75		
	1,00E+00	38,5 ± 1,2		
linear regression parameters	efficiency	1,959		
	slope	-3,425		
	Y intercept	39,6		
	DNA copy numbers tested	Average Cp value ±SD for <i>Lactobacillus</i> sp		
		<i>L.lactophilus</i>	<i>L.sakei</i>	<i>L.casei</i>
10-fold serial dilutions of DNAs	1,00E+06	19,01 ± 0,12	20,5 ± 0,12	17,79 ± 0,13
	1,00E+05	22,45 ± 0,13	23,89 ± 0,22	21,45 ± 0,1
	1,00E+04	25,82 ± 0,15	27,31 ± 0,04	24,8 ± 0,05
	1,00E+03	29,14 ± 0,1	30,36 ± 0,34	27,88 ± 0,21
	1,00E+02	32,47 ± 0,44	33,76 ± 1,04	31,33 ± 0,16
	1,00E+01	36,43 ± 0,73	35,43 ± 1,17	33,39 ± 0,83
	1,00E+00	38,32 ± 0,2	38,3 (1/3)	ND
linear regression parameters	efficiency	1,99	1,983	2,05
	slope	-3,346	-3,364	-3,2
	Y intercept	39,18	40,7	36,95
	DNA copy numbers tested	Average Cp value ±SD for <i>Lactococcus</i> sp		
		<i>L.raffinococcus</i>	<i>L.lactis</i>	<i>L.lactis cremoris</i>
10-fold serial dilutions of DNAs	1,00E+06	16,9 ± 0,12	18,18 ± 0,01	18,26 ± 0,09
	1,00E+05	20,25 ± 0,07	21,54 ± 0,13	21,75 ± 0,05
	1,00E+04	23,67 ± 0,07	24,81 ± 0,12	25,23 ± 0,13
	1,00E+03	26,77 ± 0,05	28,18 ± 0,11	27,81 ± 0,1
	1,00E+02	30,13 ± 0,11	31,47 ± 0,36	32,33 ± 0,07
	1,00E+01	33 ± 0,09	35 ± 0,19	34,49 ± 0,4
	1,00E+00	ND	ND	ND
linear regression parameters	efficiency	2	2	1,954
	slope	-3,299	-3,321	-3,438
	Y intercept	36,7	38,12	38,83

(B)

close bacterail species tested	Average Cp value $\pm$ SD						
	<i>Brochotrix</i> sp	<i>Carnobacterium</i> sp	<i>Leuconostoc</i> sp	<i>Photobacterium</i> sp	<i>Lactobacillus</i> sp	<i>Lactococcus</i> sp	<i>Pseudomonas</i> sp
<i>C.maltaromaticum</i>	ND	<b>17,78<math>\pm</math>0,45</b>	ND	ND	ND	ND	ND
contrôle négatif	ND	ND	ND	ND	ND	ND	ND
Blanc extraction	ND	ND	ND	ND	ND	ND	ND
<i>Leuconostoc gelidum</i>	ND	ND	<b>17,52 <math>\pm</math> 0,02</b>	ND	ND	ND	ND
<i>Pseudomonas aeruginosa</i>	ND	ND	ND	ND	ND	ND	<b>19,71 <math>\pm</math> 0,01</b>
<i>Pseudomonas fragi</i>	ND	ND	ND	ND	ND	ND	<b>20,5 <math>\pm</math> 0,03</b>
<b><i>Brochotrix thermophacta</i></b>	<b>19,74 <math>\pm</math> 0,02</b>	ND	ND	ND	ND	ND	ND
<i>Photobacterium angustum</i>	ND	ND	ND	<b>19,5 <math>\pm</math> 0,22</b>	ND	ND	ND
<i>Lactobacillus sakei</i>	ND	ND	ND	ND	<b>21,8<math>\pm</math> 0,03</b>	ND	ND
<i>Lactobacillus acidophilus</i>	ND	ND	ND	ND	<b>19<math>\pm</math> 0,02</b>	ND	ND
<i>Lactobacillus greminis</i>	ND	ND	ND	ND	<b>20,06 <math>\pm</math> 0,09</b>	ND	ND
<i>Lactococcus raffinolactis</i>	ND	ND	ND	ND	ND	<b>16,2 <math>\pm</math> 0,09</b>	ND
<i>Lactococcus lactis cremoris</i>	ND	ND	ND	ND	ND	<b>18,26 <math>\pm</math> 0,09</b>	ND
<i>Lactococcus piscium</i>	ND	ND	ND	ND	ND	ND	ND
<i>Leuconostoc citrium</i>	ND	ND	<b>18,52 <math>\pm</math> 0,05</b>	ND	ND	ND	ND
<i>Lactobacillus salivarius</i>	ND	ND	ND	ND	<b>18,33<math>\pm</math> 0,31</b>	ND	ND
<i>Lactococcus lactis lactis</i>	ND	ND	ND	ND	ND	<b>20,86 <math>\pm</math> 0,1</b>	ND
<i>Leuconostoc mesenteroides</i>	ND	ND	<b>18,05 <math>\pm</math> 0,35</b>	ND	ND	ND	ND
<i>lactococcus oligofermentans</i>	ND	ND	ND	ND	ND	ND	ND
<i>Serratia</i> sp	ND	ND	ND	ND	ND	ND	ND
<i>Enterobacter</i> sp	ND	ND	ND	ND	ND	ND	ND
<i>Pseudomonas azotoformans</i>	ND	ND	ND	ND	ND	ND	<b>20,14 <math>\pm</math> 0,16</b>
<i>Pseudomonas veronii</i>	ND	ND	ND	ND	ND	ND	<b>19,55 <math>\pm</math> 0,27</b>
<i>Pseudomonas poae</i>	ND	ND	ND	ND	ND	ND	<b>19,75 <math>\pm</math> 0,16</b>
<i>Pseudomonas deceptionensis</i>	ND	ND	ND	ND	ND	ND	<b>19,37 <math>\pm</math> 0,26</b>
<i>Pseudomonas gessardii</i>	ND	ND	ND	ND	ND	ND	<b>20 <math>\pm</math> 0,09</b>

## 5 Correlation between classic microbiology, targeted metagenomic and quantitative data : example on on real foodstuff samples.

6 naturally contaminated matrices from different origins (white pudding, pancake, hamburger, ham, pasta, cheese) were analyzed by microbiology, metagenomic analysis and quantitative PCR to estimate proportion of some major taxons (*Carnobacterium* sp, *Leuconostoc* sp, *Photobacterium* sp, *Brochotrix* sp, *Pseudomonas* sp, *Lactococcus* sp et *Lactobacillus* sp). Among the 7 taxons tests in qPCR, 3 are detected in all the 6 food samples (*Leuconostoc* sp, *Lactococcus* sp et *Lactobacillus* sp). Thus, the total bacterial flora could be deduced using these qPCR data. The total flora estimated by classic microbiology is always widely mis-estimated compared to total flora deduced by qPCR

(fig 19). The qPCR method allow to estimate more accurately total bacterial flora (including cultivated or not cultivated bacteria).

5 These deduction of total flora by qPCR method will allow to quantify more precisely all the taxons identified by metagenomic analysis (fig 20). The figure 20 (A) shows the bacterial profile obtained by metagenomic analysis for the 6 food samples. These analyses don't allow estimation of the concentration of each taxon denominated in colony-forming units per gram of matrix (cfu/g). The total flora estimation by quantitative PCR combining with the percentage of reads from genomic analysis allow to  
10 obtain an estimation of each taxon concentration in colony-forming unit per gram of matrix (fig 20 (B)).

The invention is not limited by the examples or figures provided herein. The skilled person is able to conceive of alternatives for the various sequences, tools and methods  
15 mentioned in this specification. The invention is limited only by the scope of the claims.

**Claims**

1. A method of metagenomic analysis of a food sample to detect and quantitate  
5 microorganisms present in said sample, the analysis comprising amplification of  
the V2-V3 region of bacterial 16S rRNA in said sample, wherein the method  
further comprises the use of Peptide Nucleic Acid (PNA) clamping to reduce  
amplification of at least one of: i) chloroplast DNA present in the sample; ii)  
10 mitochondrial DNA present in the sample; and/or iii) DNA from one or more  
selected microorganisms that may be present in the sample and wherein  
propidium monoazide (PMA) is added to the sample prior to DNA extraction in  
order to reduce the subsequent amplification of DNA from non-viable  
microorganisms present in the sample.
- 15
2. A method of claim 1 wherein the V1-V3 region of bacterial 16S rRNA is  
amplified.
- 20
3. A method of claim 1 or claim 2 wherein said analysis can detect and distinguish  
bacteria of the families: *Acetobacteraceae*, *Aeromonadaceae*, *Bacteroidaceae*,  
*Burkholderiaceae*, *Carnobacteriaceae*, *Clostridiaceae*, *Comamonadaceae*,  
*Enterobacteriaceae*, *Enterococcaceae*, *Flavobacteriaceae*, *Lachnospiraceae*,  
*Lactobacillaceae*, *Leuconostocaceae*, *Listeriaceae*, *Moraxellaceae*, *Neisseriaceae*,  
25 *Pseudoalteromonadaceae*, *Pseudomonadaceae*, *Shewanellaceae*,  
*Streptococcaceae*, *Vibrionaceae* and *Xanthomonadaecea*.
4. A method of any preceding claim wherein said analysis can detect and distinguish  
bacteria of the genres: *Acetobacteraceae* *Acetobacter*, *Acetobacteraceae*  
30 *Gluconacetobacter*, *Aeromonadaceae* *Aeromonas*, *Bacteroidaceae* *Bacteroides*,  
*Burkholderiaceae* *Burkholderia*, *Carnobacteriaceae* *Carnobacterium*,  
*Clostridiaceae* *Clostridium*, *Comamonadaceae* *Comamonas*, *Comamonadaceae*  
*Delftia*, *Enterobacteriaceae* *Escherichia*, *Enterobacteriaceae* *Pantoea*,  
*Enterobacteriaceae* *Serratia*, *Enterococcaceae* *Enterococcus*, *Chryseobacterium*,  
35 *Flavobacteriaceae* *Flavobacterium*, *Flavobacteriaceae* *Myroides*,

*Lachnospiraceae Butyrivibrio, Lactobacillaceae Lactobacillus, Lactobacillaceae Pediococcus, Leuconostocaceae Leuconostoc, Leuconostocaceae Weissella, Listeriaceae Listeria, Moraxellaceae Acinetobacter, Moraxellaceae Psychrobacter, Neisseriaceae Alysella, Pseudoalteromonadaceae Pseudoalteromonas, Pseudomonadaceae Pseudomonas, Shewanellaceae Shewanella, Streptococcaceae Lactococcus, Streptococcaceae Streptococcus, Vibrionaceae Aliivibrio, Vibrionaceae Photobacterium, Vibrionaceae Vibrio and Xanthomonadaecea Xanthomonas.*

- 5
- 10
- 15
- 20
- 25
- 30
5. A method of any preceding claim wherein the percentage of microorganisms represented by at least one operational taxonomic unit (OTU) is determined.
  6. A method of claim 5 further comprising determining the colony forming units count (CFU) of the sample and normalising the percentage of organisms represented by said OTU to a proportion of the total viable CFU count of the sample.
  7. A method of any preceding claim wherein the said metagenomic analysis is followed by taxa specific qPCR.
  8. A method of any preceding claim wherein the metagenomic analysis is conducted using a primer comprising the sequence of SEQ ID NO. 1.
  9. A method of any preceding claim wherein the metagenomic analysis is conducted using a primer comprising the sequence of SEQ ID NO. 2.
  10. A method of any preceding claim wherein the metagenomic analysis is conducted using a pair of primers comprising the sequences of SEQ ID NO 1 and SEQ ID NO. 2.
  11. A method of any of claims 8 to 10 wherein said primers consist of the sequences of SEQ ID NO. 1 or SEQ ID NO. 2 respectively.

12. A method of any of claims 8 to 11 wherein said primer is modified by the addition of an adaptor, key sequence, tag or universal sequence, optionally having the sequences of SEQ ID NO.s 82 and 83 respectively.

5

13. A method of claim 1 wherein said PNA clamping comprises conducting metagenomic analysis in the presence of a PNA oligonucleotide comprises a sequence derived from a chloroplast 16S rRNA gene, preferably a sequence common to a plurality of chloroplast 16S rRNA genes.

10

14. A method of claim 13 wherein said PNA oligonucleotide comprises a sequence common to at least two of, and preferably all of, the chloroplast 16S rRNA sequences of SEQ ID NOs 3 to 10, or common to at least two of, and preferably all of, the chloroplast 16S rRNA sequences of SEQ ID NOs 18 to 25.

15

15. A method of claim 13 or claim 14 wherein said PNA oligonucleotide has a sequence not found in the bacterial 16S rRNA sequences of at least one of, and preferably all of, SEQ ID NOs 11 to 16, or has a sequence not found in the bacterial 16S rRNA sequences of at least one of, and preferably all of, SEQ ID NOs 26 to 31.

20

16. A method of any of claims 13 to 15 wherein said PNA oligonucleotide comprises the sequence of SEQ ID NO. 17, optionally wherein said PNA oligonucleotide consists of the sequence of SEQ ID NO. 17.

25

17. A method of any of claims 13 to 15 wherein said PNA oligonucleotide comprises the sequence of SEQ ID NO. 32, optionally wherein said PNA oligonucleotide consists of the sequence of SEQ ID NO. 32.

30

18. A method of claim 1 wherein said PNA clamping comprises conducting metagenomic analysis in the presence of a PNA oligonucleotide comprises a sequence derived from a mitochondrial 18S rRNA gene, preferably a sequence common to a plurality of mitochondrial 18S rRNA genes.

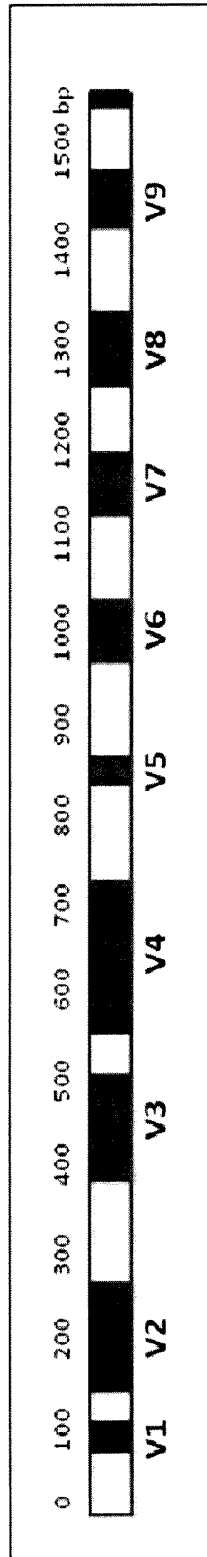
19. A method of claim 18 wherein said PNA oligonucleotide comprises a sequence common to at least two of, and preferably all of, the mitochondrial 18S rRNA sequences of SEQ ID NOs 33 to 38.
- 5 20. A method of claim 18 or claim 19 wherein said PNA oligonucleotide has a sequence not found in the chloroplast 16S rRNA sequences, or the bacterial 16S rRNA sequences of at least one of, and preferably all of, SEQ ID NOs 39 to 47.
- 10 21. A method of any of claims 18 to 20 wherein said PNA oligonucleotide comprises the sequence of SEQ ID NO. 48, optionally wherein said PNA oligonucleotide consists of the sequence of SEQ ID NO. 48.
- 15 22. A method of claim 1 wherein said PNA clamping comprises conducting metagenomic analysis in the presence of a PNA oligonucleotide comprises a sequence derived from a bacterial 16S rRNA gene, preferably a sequence common to a plurality of bacterial 16S rRNA genes.
- 20 23. A method of claim 22 wherein said PNA oligonucleotide comprises a sequence common to at least two of, and preferably all of, the bacterial 16S rRNA sequences of SEQ ID NOs 49 to 61.
- 25 24. A method of claim 22 or claim 23 wherein said PNA oligonucleotide has a sequence not found in the bacterial 16S rRNA sequences of at least one of, and preferably all of, SEQ ID NOs 62 to 79.
- 30 25. A method of any of claims 22 to 24 wherein said PNA oligonucleotide comprises the sequence of SEQ ID NO. 80, optionally wherein said PNA oligonucleotide consists of the sequence of SEQ ID NO. 80
- 35 26. A kit for metagenomic analysis of a sample comprising one or both primers comprising the sequence of SEQ ID NO. 1 and/or SEQ ID NO. 2, preferably wherein the primers consist of the sequences of SEQ ID NO. 1 and/or SEQ ID NO. 2 and further comprising a PNA oligonucleotide comprising a sequence of



any of SEQ ID NOs. 17, 32, 48 or 80, preferably wherein the primer consists of the sequence of any of SEQ ID NOs. 17, 32, 48 or 80.

- 5 27. A kit of claim 26 wherein the primer(s) are modified by the addition of an adaptor, key sequence, tag or universal sequence, optionally having the sequences of SEQ ID NO.s 82 and 83 respectively.
- 10 28. A kit of claim 26 or 27 comprising a set of primers including any two, three or four PNA oligonucleotides comprising or consisting of the sequence of SEQ ID NO.s 17, 32, 48 or 80.
- 15 29. A kit of any of claims 26 to 28 further comprising propidium monoazide (PMA).
30. A kit of any of claims 26 to 29 further comprising a reagent for metagenomic analysis of a sample, e.g. a food sample.

**FIGURE 1**



5

10

FIGURE 2

Percentage of species with differences in hypervariable regions 16s rDNA

Bacterial Genres	VARIABLE REGIONS														
	V2	V2V3	V3	V3V4	V4	V4V5	V5	V5V6	V6	V6V7	V7	V7V8	V8	V8V9	V9
Acetobacteraceae_Acetobacter	77%	83%	30%	92%	87%	90%	55%	72%	61%	70%	38%	80%	60%	75%	26%
Acetobacteraceae_Gluconacetobacter	63%	81%	25%	88%	81%	81%	47%	59%	41%	63%	38%	59%	35%	65%	25%
Aeromonadaceae_Aeromonas	47%	83%	75%	78%	28%	31%	14%	42%	26%	34%	14%	19%	22%	38%	17%
Bacteroidaceae_Bacteroides	98%	99%	93%	97%	96%	88%	81%	87%	95%	98%	87%	96%	92%	100%	87%
Burkholderiaceae_Burkholderia	77%	100%	89%	95%	48%	67%	23%	79%	66%	75%	18%	48%	33%	55%	23%
Carnobacteriaceae_Carnobacterium	100%	100%	60%	80%	70%	90%	80%	100%	90%	100%	50%	90%	89%	100%	75%
Clostridiaceae_Clostridium	94%	96%	77%	95%	88%	92%	72%	90%	85%	91%	69%	92%	87%	97%	93%
Comamonadaceae_Comamonas	100%	100%	100%	100%	73%	100%	80%	100%	92%	92%	42%	67%	30%	80%	83%
Comamonadaceae_Delftia	33%	67%	50%	50%	67%	67%	67%	100%	67%	100%	67%	67%	67%	100%	100%
Enterobacteriaceae_Escherichia	100%	100%	83%	100%	83%	100%	83%	83%	67%	67%	50%	67%	67%	67%	33%
Enterobacteriaceae_Pantoea	58%	90%	80%	95%	50%	80%	45%	85%	60%	80%	30%	53%	58%	73%	36%
Enterobacteriaceae_Serratia	53%	93%	80%	93%	47%	53%	27%	53%	40%	73%	47%	80%	87%	91%	18%
Enterococcaceae_Enterococcus	69%	86%	58%	81%	48%	55%	23%	59%	45%	59%	28%	55%	40%	73%	98%
Flavobacteriaceae_Chryseobacterium	100%	100%	65%	93%	72%	88%	64%	86%	47%	71%	21%	66%	36%	77%	32%
Flavobacteriaceae_Flavobacterium	100%	100%	90%	90%	88%	94%	60%	86%	60%	84%	56%	82%	69%	81%	90%
Flavobacteriaceae_Myroides	80%	100%	100%	100%	100%	100%	100%	100%	50%	100%	80%	100%	40%	80%	100%
Lachnospiraceae_Butyrvibrio	*	100%	100%	100%	100%	100%	67%	100%	100%	100%	67%	100%	100%	100%	100%
Lactobacillaceae_Lactobacillus	85%	93%	70%	89%	75%	80%	40%	76%	65%	78%	41%	82%	64%	90%	98%
Lactobacillaceae_Pediococcus	87%	100%	73%	100%	87%	100%	54%	85%	73%	93%	64%	93%	80%	75%	50%
Leuconostocaceae_Leuconostoc	89%	95%	90%	90%	70%	80%	45%	61%	55%	70%	50%	74%	63%	80%	100%
Leuconostocaceae>Weissella	80%	93%	67%	93%	71%	86%	60%	85%	69%	93%	62%	92%	71%	75%	100%
Listeriaceae_Listeria	90%	90%	44%	44%	50%	50%	20%	33%	22%	50%	40%	100%	90%	100%	33%
Moraxellaceae_Acinetobacter	81%	96%	96%	96%	72%	80%	44%	57%	48%	58%	37%	63%	20%	100%	80%
Moraxellaceae_Psychrobacter	100%	75%	66%	94%	67%	67%	13%	27%	16%	35%	30%	39%	27%	71%	76%
Neisseriaceae_Alysiella	100%	100%	100%	100%	50%	100%	100%	100%	100%	100%	100%	100%	50%	*	*
Pseudoalteromonadaceae_Pseudoalterom.	50%	79%	58%	74%	39%	70%	54%	81%	49%	63%	22%	56%	61%	79%	43%
Pseudomonadaceae_Pseudomonas	57%	90%	78%	89%	62%	74%	43%	65%	47%	60%	28%	60%	40%	68%	44%
Shewanellaceae_Shewanella	54%	94%	95%	96%	67%	80%	33%	71%	53%	71%	29%	67%	57%	89%	48%
Streptococcaceae_Lactococcus	100%	100%	78%	75%	50%	50%	38%	50%	38%	63%	50%	75%	63%	50%	50%
Streptococcaceae_Streptococcus	100%	100%	76%	85%	74%	78%	39%	84%	68%	82%	28%	69%	45%	82%	69%
Vibrionaceae_Aliivibrio	50%	100%	100%	100%	50%	67%	50%	100%	100%	100%	17%	50%	50%	100%	100%
Vibrionaceae_Photorbacterium	57%	96%	96%	96%	57%	74%	30%	83%	43%	65%	35%	78%	74%	80%	30%
Vibrionaceae_Vibrio	60%	96%	92%	97%	60%	66%	24%	61%	48%	60%	18%	63%	48%	94%	56%
Xanthomonadaceae_Xanthomonas	30%	44%	37%	67%	56%	56%	15%	30%	22%	33%	22%	44%	30%	38%	19%

## FIGURE 3

SEQ ID NO. 81

5        01    GaGaGtT **TGaT**Catgget **CAG**ATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAAC

61        **GGTAACAGGAAGCAGCTTGCTTCTTTGCTGACG**AGTggcGGaCgGgTgAGTAAgTCTGG    **V1**

121       GAgACgGCg**TGATGGAGGGGATAACTACTGAAACGGTAGCTAATACCGCATAACGTCG**    **V2**

10       181       **CAAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCG**GATGTGCCCAGATGGGATTAG

241       CTAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGAC

15       301       CAGCCACACTGGAActGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATAT

361       TGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTT

421       GTAAAGTACTTT**CAGCGGGGAGGAAGGGAGTAAAGTAAATACCTTTGCTCATTGACGTTA**    **V3**

20       481       **CCCGCAGAAG**AAGCACCGGCTaaCtCCgtgcccagcagccggtaaTacGGAGGGTGCAA

541       GCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGCGG**TTTGTTAAGTCAGATGTGAA**    **V4**

25       601       **ATCCCCGGGCTCAACCTGGGAActGCATCTGATACTGGCAAGCT**TGAGTCTCGTAGAGGG

661       GGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGA

721       AGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGAT

30       781       TAGATACCCTGGTAGTCCACGCCGTAAACGATGTGCACTT**GGAGGTTGTGCCCTTGAGGC**    **V5**

841       **GTGGCTTCC**GGAGCTAACCGGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAA

35       901       ACTCAAATGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTAATTCGATGCA

40

5

## FIGURE 3 cont.

961 ACGCGAAGAACCTTACCTGGTCTTGACATCC **ACGGAAGTTTTCAGAGATGAGAATGTGCC V6**  
1021 **TTCGGGAAC**CGTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTGTGAAATGTTGG  
1081 GTTAAGTCCCGCAACGAGCGCAACCCTTAT **CCTTTGTTGCCAGCGGTCCGGCCGGGAAC V7**  
1141 **CAAAGGAGA**CTGCCAGTGATAAACTGGAGGAAGGTGGGATGACGTCAAGTCATCATGGC  
1201 CCTTACGACCAGGGCTACACACGTGCTACAATGG **CGCATACAAAGAGAAGCGACCTCGCG V8**  
1261 **AGAGCAAGCGGACCTCATAAAGTGCGT**CGTAGTCCGGATTGGAGTCTGCAACTCGACTCC  
1321 ATGAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGTTCCCGGGC  
1381 CTTGTACACACCGCCCGTACACCATGGGAGTGGGTTGCAAAAGAA **GTAGGTAGCTTAAC V9**  
1441 **CTTCGGGAGGGCGCTA**CCACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAA

**TOTALLY CONSERVED** conserved(lowercase) Variable(capital) **Hypervariable**

## FIGURE 4

5

V1-V3

10	OLIGO	start	len	tm	gc%
	<b>16S-V1-V3-FW</b>	1	21	57.47	47.62
	GAGAGTTTGATYMTGGCTCAG				
15	SEQ ID NO. 1				
	<b>16S-V1-V3-Rev</b>	555	17	69.56	76.47
	ACCGCGGCTGCTGGCAC				
20	SEQ ID NO. 2				

PRODUCT SIZE: 555, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 2.00

FIGURE 5

5

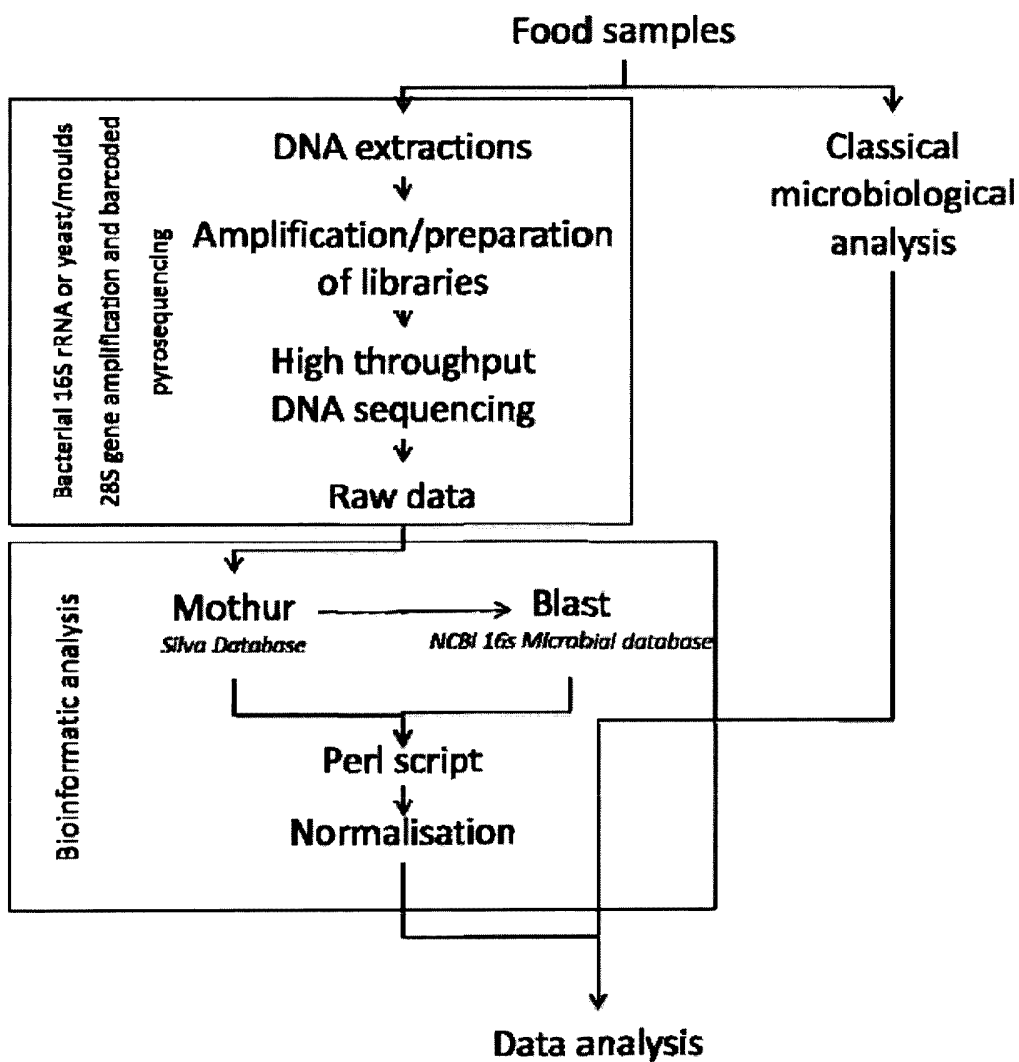


FIGURE 6

5

CHLOROPLAST1 GAGAGTTGGATCCCTGGCTCAGGATGAACGGCTGGCGCATGCTTAACACATG-CAACTCCG-----ACGGG-----AAGTGGT--GTT-----TCCAGTGG  
CHLOROPLAST2 GAGAGTTGGATCCCTGGCTCAGGATGAACGGCTGGCGCATGCTTAACACATG-CAACTCCG-----ACGGG-----AAGTGGT--GTT-----TCCAGTGG  
CHLOROPLAST3 GAGAGTTGGATCCCTGGCTCAGGATGAACGGCTGGCGCATGCTTAACACATG-CAACTCCG-----ACGGG-----AAGTGGT--GTT-----TCCAGTGG  
CHLOROPLAST4 GAGAGTTGGATCCCTGGCTCAGGATGAACGGCTGGCGCATGCTTAACACATG-CAACTCCG-----ACGGG-----AAGTGGT--GTT-----TCCAGTGG  
CHLOROPLAST5 GAGAGTTGGATCCCTGGCTCAGGATGAACGGCTGGCGCATGCTTAACACATG-CAACTCCG-----ACGGG-----AAGTGGT--GTT-----TCCAGTGG  
CHLOROPLAST6 GAGAGTTGGATCCCTGGCTCAGGATGAACGGCTGGCGCATGCTTAACACATG-CAACTCCG-----ACGGG-----AAGTGGT--GTT-----TCCAGTGG  
CHLOROPLAST7 GAGAGTTGGATCCCTGGCTCAGGATGAACGGCTGGCGCATGCTTAACACATG-CAACTCCG-----ACGGG-----AAGTGGT--GTT-----TCCAGTGG  
CHLOROPLAST8 GAGAGTTGGATCCCTGGCTCAGGATGAACGGCTGGCGCATGCTTAACACATG-CAACTCCG-----ACGGG-----AAGTGGT--GTT-----TCCAGTGG  
  
Pseudomonas GAGAGTTGGATCCCTGGCTCAGGATGAACGGCTGGCGCATGCTTAACACATG-CAACTCCG-----ACGGG-----AAGTGGT--GTT-----TCCAGTGG  
Xanthomonas GAGAGTTGGATCCCTGGCTCAGGATGAACGGCTGGCGCATGCTTAACACATG-CAACTCCG-----ACGGG-----AAGTGGT--GTT-----TCCAGTGG  
E. Coli GAGAGTTGGATCCCTGGCTCAGGATGAACGGCTGGCGCATGCTTAACACATG-CAACTCCG-----ACGGG-----AAGTGGT--GTT-----TCCAGTGG  
Lactococcus GAGAGTTGGATCCCTGGCTCAGGATGAACGGCTGGCGCATGCTTAACACATG-CAACTCCG-----ACGGG-----AAGTGGT--GTT-----TCCAGTGG  
Brochothrix GAGAGTTGGATCCCTGGCTCAGGATGAACGGCTGGCGCATGCTTAACACATG-CAACTCCG-----ACGGG-----AAGTGGT--GTT-----TCCAGTGG  
Arthrobacter GAGAGTTGGATCCCTGGCTCAGGATGAACGGCTGGCGCATGCTTAACACATG-CAACTCCG-----ACGGG-----AAGTGGT--GTT-----TCCAGTGG  
**QP-MetaFv GAGAGTTGGATCCCTGGCTCAGGATGAACGGCTGGCGCATGCTTAACACATG-CAACTCCG-----ACGGG-----AAGTGGT--GTT-----TCCAGTGG**

15

20



FIGURE 7

5

CHLOROPLAST1 GGAGAGAAG-----CAATGACGGTATCTGGGGAATAAGCATCGGCTAACTCTGTGTGCCACGACGCCCGCGGT  
 CHLOROPLAST2 GGAGAGAAG-----CAATGACGGTATCTGGGGAATAAGCATCGGCTAACTCTGTGTGCCACGACGCCCGCGGT  
 CHLOROPLAST3 GGAGAGAAG-----CAATGACGGTATCTGGGGAATAAGCATCGGCTAACTCTGTGTGCCACGACGCCCGCGGT  
 CHLOROPLAST4 GGAGAGAAG-----CAATGACGGTATCTGGGGAATAAGCATCGGCTAACTCTGTGTGCCACGACGCCCGCGGT  
 CHLOROPLAST5 GGAGAGAAG-----CAATGACGGTATCTGGGGAATAAGCATCGGCTAACTCTGTGTGCCACGACGCCCGCGGT  
 CHLOROPLAST6 AGAGAGAAG-----CAATGACGGTATCTGGGGAATAAGCATCGGCTAACTCTGTGTGCCACGACGCCCGCGGT  
 CHLOROPLAST7 GGAGAGAAG-----CAATGACGGTATCTGGGGAATAAGCATCGGCTAACTCTGTGTGCCACGACGCCCGCGGT  
 CHLOROPLAST8 GGAGAGAAG-----CAATGACGGTATCTGGGGAATAAGCATCGGCTAACTCTGTGTGCCACGACGCCCGCGGT

AR-----CAATGACGGTATCTG **QP-PNAcdn**

Pseudomonas sp TGGGAGGAAGGCGAGTAAGCGGAATACCTTGCCTTTTGACGTTACCGACAGAAATAAGCACCGGCTAACTCTGTGTGCCACGACGCCCGCGGT  
 Xanthomonas sp GGGGAAGAAATCCTGCTGGCTAATAATCCGGCGGGATGACCGTACCCAAAGAAATAACACCCGGCTAACTCTGTGTGCCACGACGCCCGCGGT  
 E. Coli sp GGGGAGGAAGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACCCGCGAGAAAGACCCGGCTAACTCTGTGTGCCACGACGCCCGCGGT  
 Lactococcus sp AGAGAGAAGTCCGTGAGAGTAACTGTTACGTTTCGACGGTATCTAACCCAGAAAGCCACGGCTAACTCTGTGTGCCACGACGCCCGCGGT  
 Brochothrix sp AGAGAGAACAATGGGTGAGAGTAACTGTTACCCNCTTACCGGTTAATACCCAGAAAGCCACGGCTAACTCTGTGTGCCACGACGCCCGCGGT  
 Arthrobacter sp AGGGTAGAA-GCCTGTCTTAGGGTTCGGT-----GACGGTACCTTCAGAGAAGGCGCGGCTAACTCTGTGTGCCACGACGCCCGCGGT

**GTGCCACGACGCCCGGT QP-MetaRev**

25

**FIGURE 8**

5

Mito-Oryza-sativa  
Mito-Oryza-rufipogon  
Mito-Bambusa-oldhamii  
Mito-Triticum-aestivum  
Mito-S.cereale  
Mito-Sorghum-bicolor  
10  
-----CTGTTTGTAGACCTCGGTAGTATTAGGTACTTGGTCGGGTAAGA  
-----CTGTTTGTAGACCTCGGTAGTATTAGGTACTTGGTCGGGTAAGA  
-----CTGTTTGTAGACCTCGGTAGTATTAGGTACTTGGTCAGGTAAGA  
-----CTGTTTGTAGACCTCGGTAGTATTAGGTACTTGGTCAGGTAAGA  
-----CTGTTTGTAGACCTCGGTAGTATTAGGTACTTGGTCAGGTAAGA  
-----CTGTTTGTAGACCTCGGTAGTATTAGGTACTTGGTCAGGTAAGA  
-----TGTGTGATGAGCCTGCGT--- **QP-PNam** -----  
-----GCCCCGAGAGGGGCTCGCGTCTGATTAGCTAGTGGTGAGGCAATA  
-----GCCCGAGAGGGGCTCGCGTCTGATTAGCTAGTGGTGAGGCAATA  
-----GCCCAAGGAGGGGCTCGCGTCTGATTAGCTAGTGGTGAGGCAATA  
-----GCCCAAGGAGGGGCTCGCGTCTGATTAGCTAGTGGTGAGGCAATA  
E.Coli sp  
Lactoc-lactis sp  
Weissella  
Brochotr-Therm sp  
Stapylococcus sp  
15  
-----TGGCATCGGATGCCCCAGATGGGATTAGCTTGTGGTGGGTAACG  
TATCACTTTGGGATGGACCTCGCGGCGCATTAGCTTGTGGTAGGGTAACG  
TATCACTAAGAGATGGTCCCGGGTGCATTAGCTAGTGGTAGGGTAATG  
TGTCACGTGAGATGGACCCCGCTGGATTAGCTAGTGGTAGGGTAATG  
TATCACTTATAGATGGACCCCGCTGGATTAGCTAGTGGTAGGGTAACG  
20

25

**FIGURE 9**

	Lactobacillus-delbrueckii-subs	GGCTACCAAGGCAATGATGCGTAGCCGAGTTGAGAGACTGATCGGGCCA
5	Lactobacillus-delbrueckii-subs	GGCTACCAAGGCAATGATGCGTAGCCGAGTTGAGAGACTGATCGGGCCA
	Lactobacillus-amylovorus	GGCTACCAAGGCAATGATGCGTAGCCGAGTTGAGAGACTGATCGGGCCA
	Lactobacillus-acidophilus	GGCTACCAAGGCAATGATGCGTAGCCGAGTTGAGAGACTGATCGGGCCA
	Lactobacillus-amylolyticus	GGCTACCAAGGCAATGATGCGTAGCCGAGTTGAGAGACTGATCGGGCCA
	Lactobacillus-acetotolerans	GGCTACCAAGGCAATGATGCGTAGCCGAGTTGAGAGACTGATCGGGCCA
10	Lactococcus-piscium	GGACTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Lactococcus-plantarum	GGACTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Lactococcus-chungangensis	GGACTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Lactococcus-raffinolactis	GGACTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Lactococcus-lactis-subsp-Hordn	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
15	Lactococcus-lactis-subsp-Lacti	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Lactococcus-lactis-subsp-Cremo	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	<b>PNALactococcus</b>	-----GATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA-----
	Streptococcus-Thermophilus-19-5	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Streptococcus-Thermophilus-UKPS2	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
20	streptococcus-Thermophilus-SM1	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Streptococcus-parauberis	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Streptococcus-iniae	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Streptococcus-dysgalactiae-ssp	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Streptococcus-agalactiae	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
25	Streptococcus-equi	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Streptococcus-oligofermentans	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Enterococcus-faecalis	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Lactobacillus-algidus	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Lactobacillus-frumenti	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
30	Lactobacillus-panis	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Lactobacillus-fermentum	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Lactobacillus-oligofermentans	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Lactobacillus_plantarum	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
	Brevibacterium-linens	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA
35	Psychroflexus-sp	GGCTACCAAGGGGATGATACATAGCCGACCTGAGAGGTTGATCGGGCCA

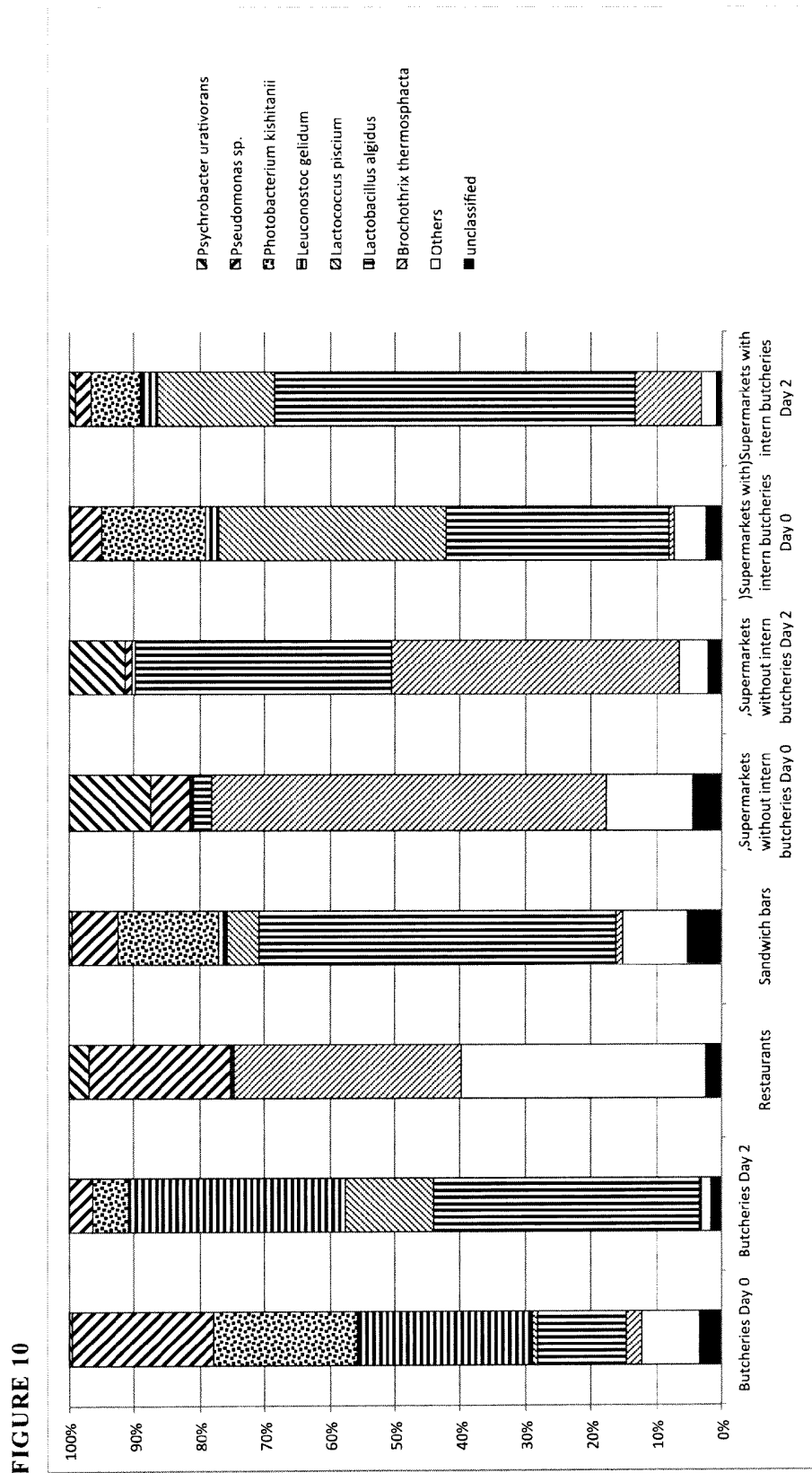
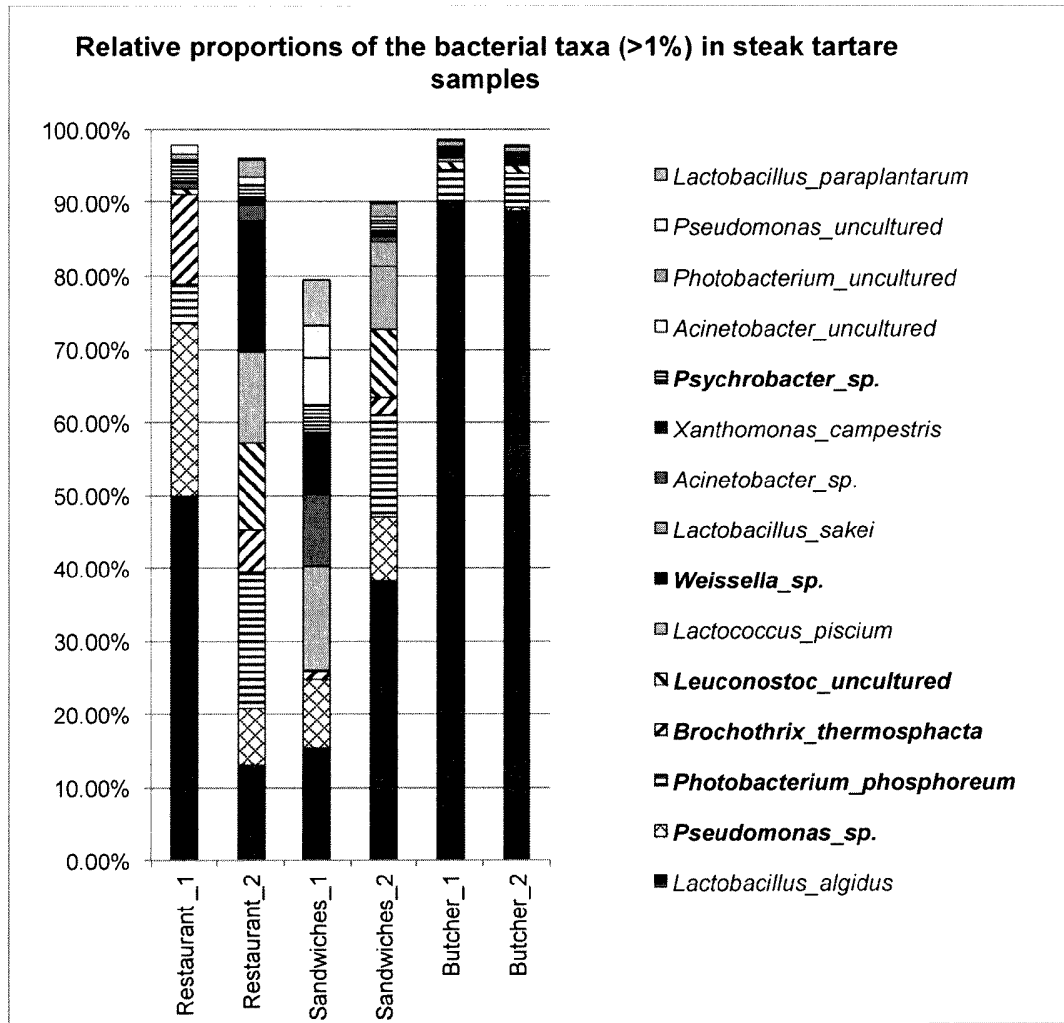
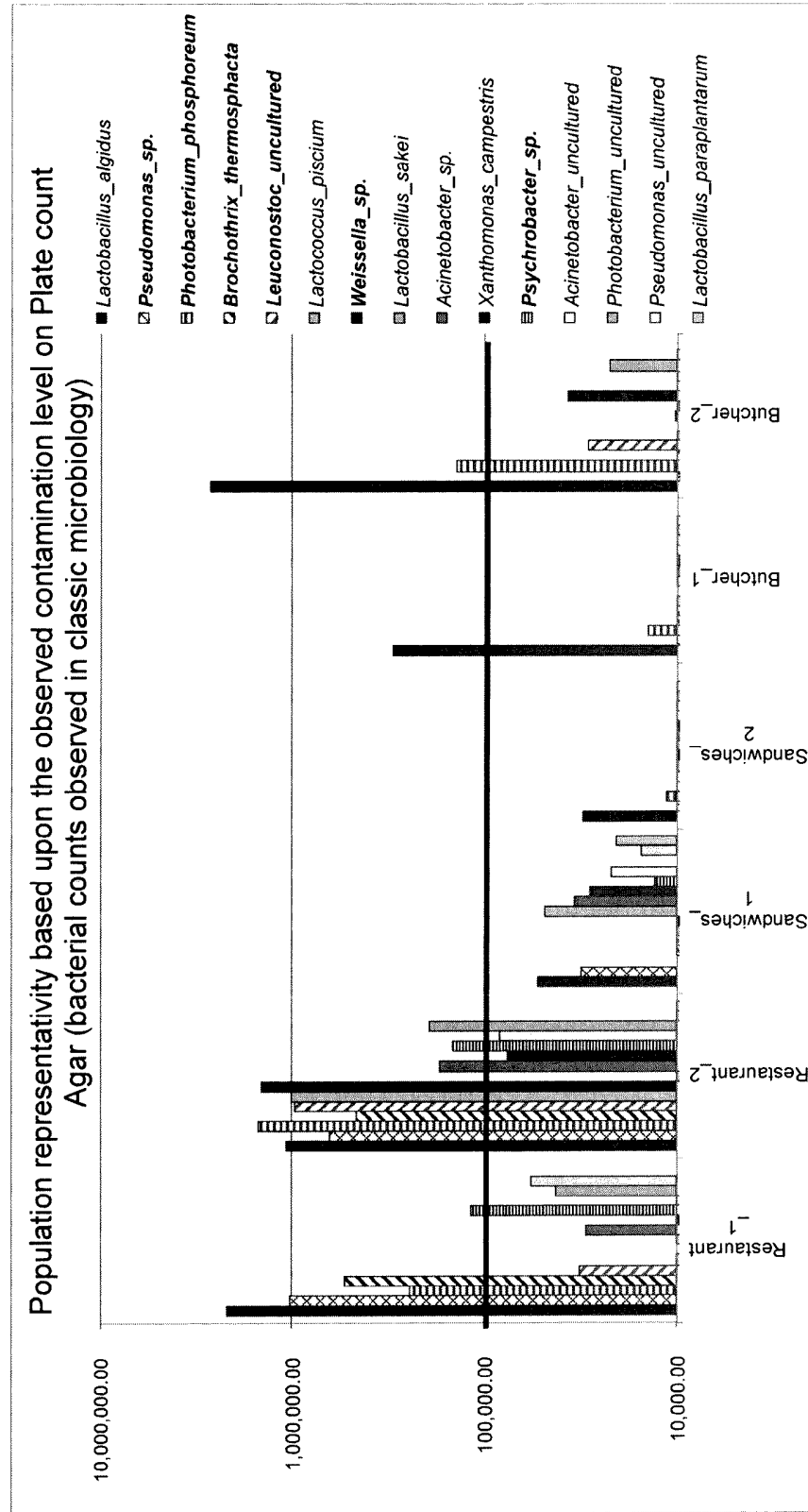


FIGURE 11



5

FIGURE 12



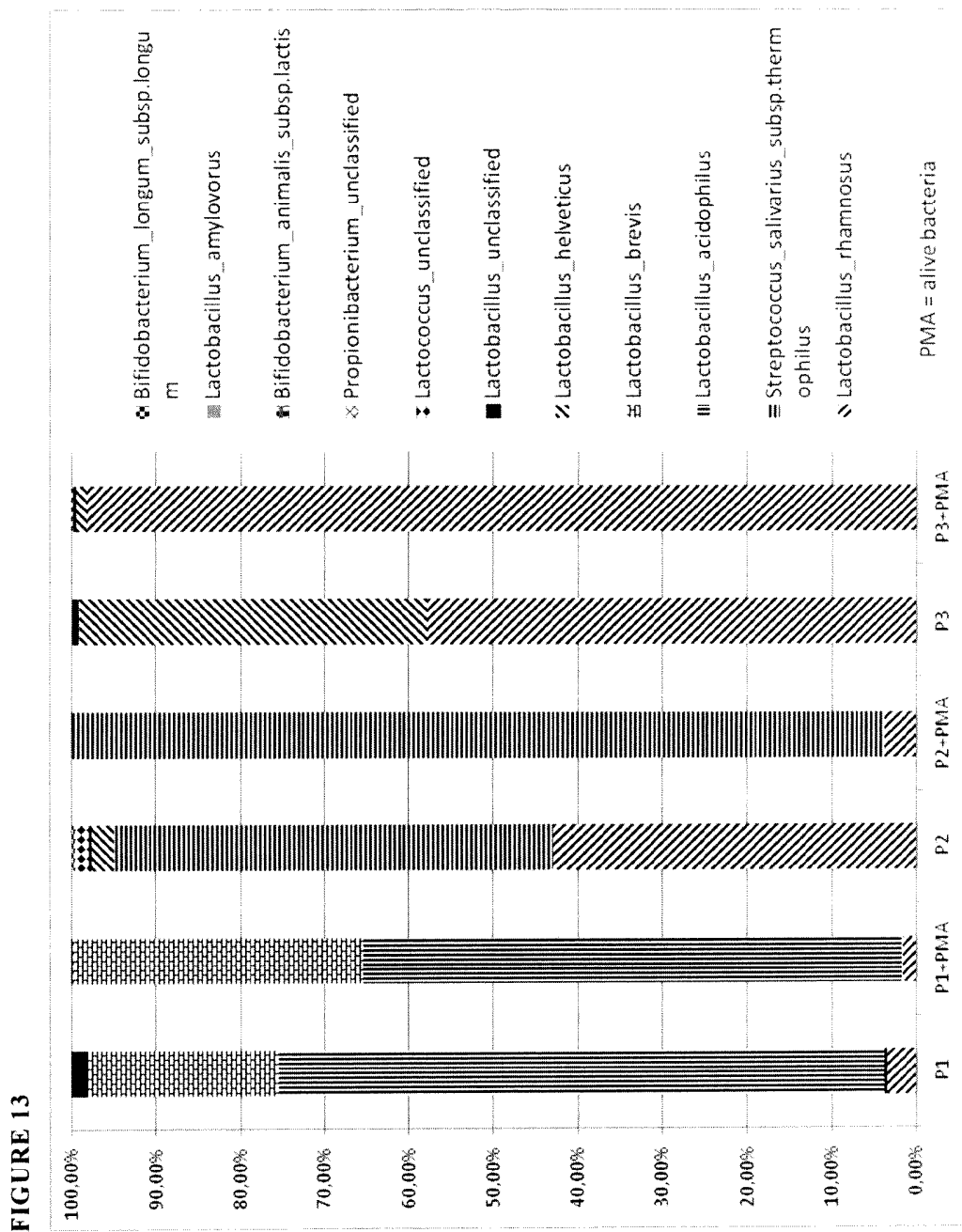
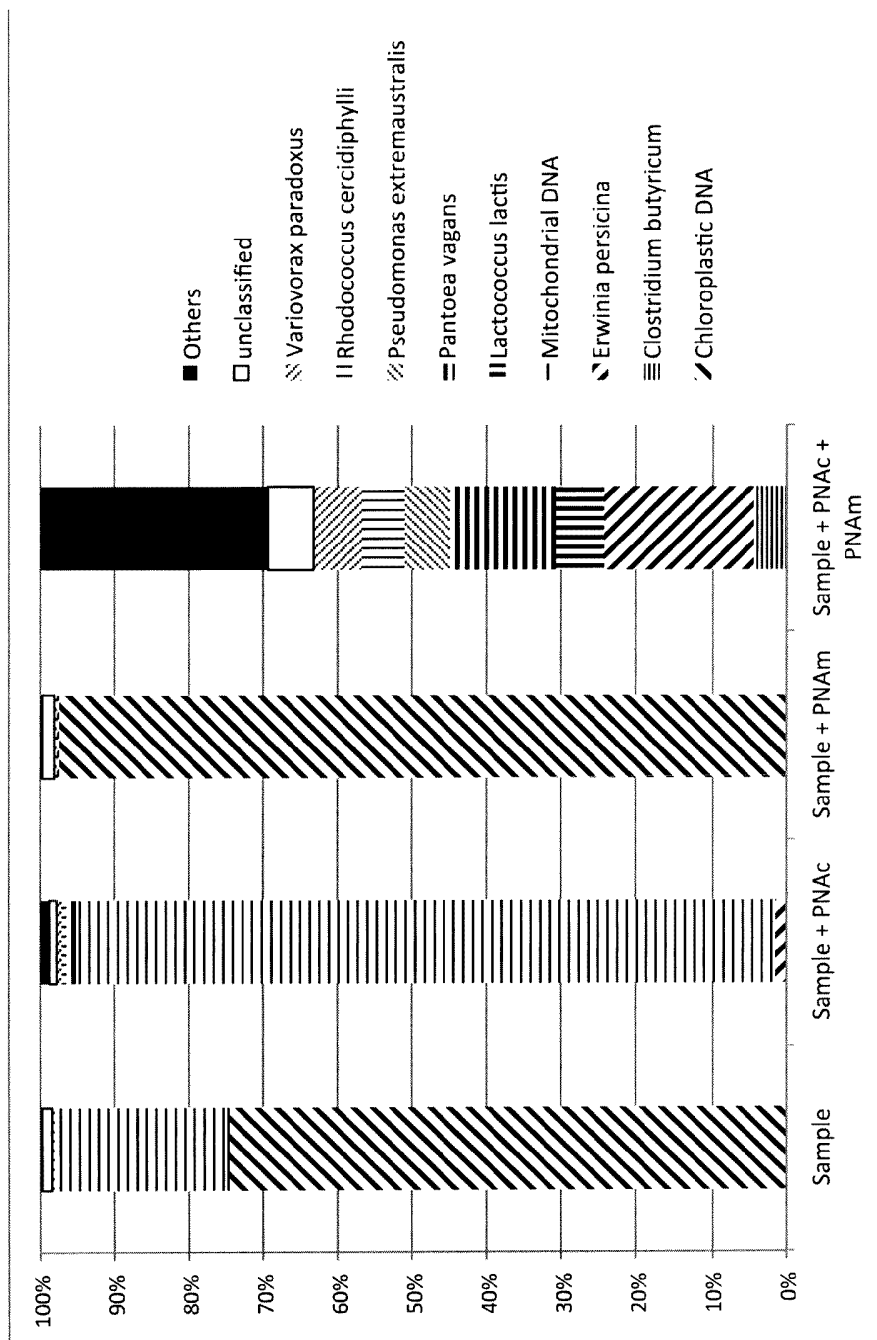


FIGURE 14





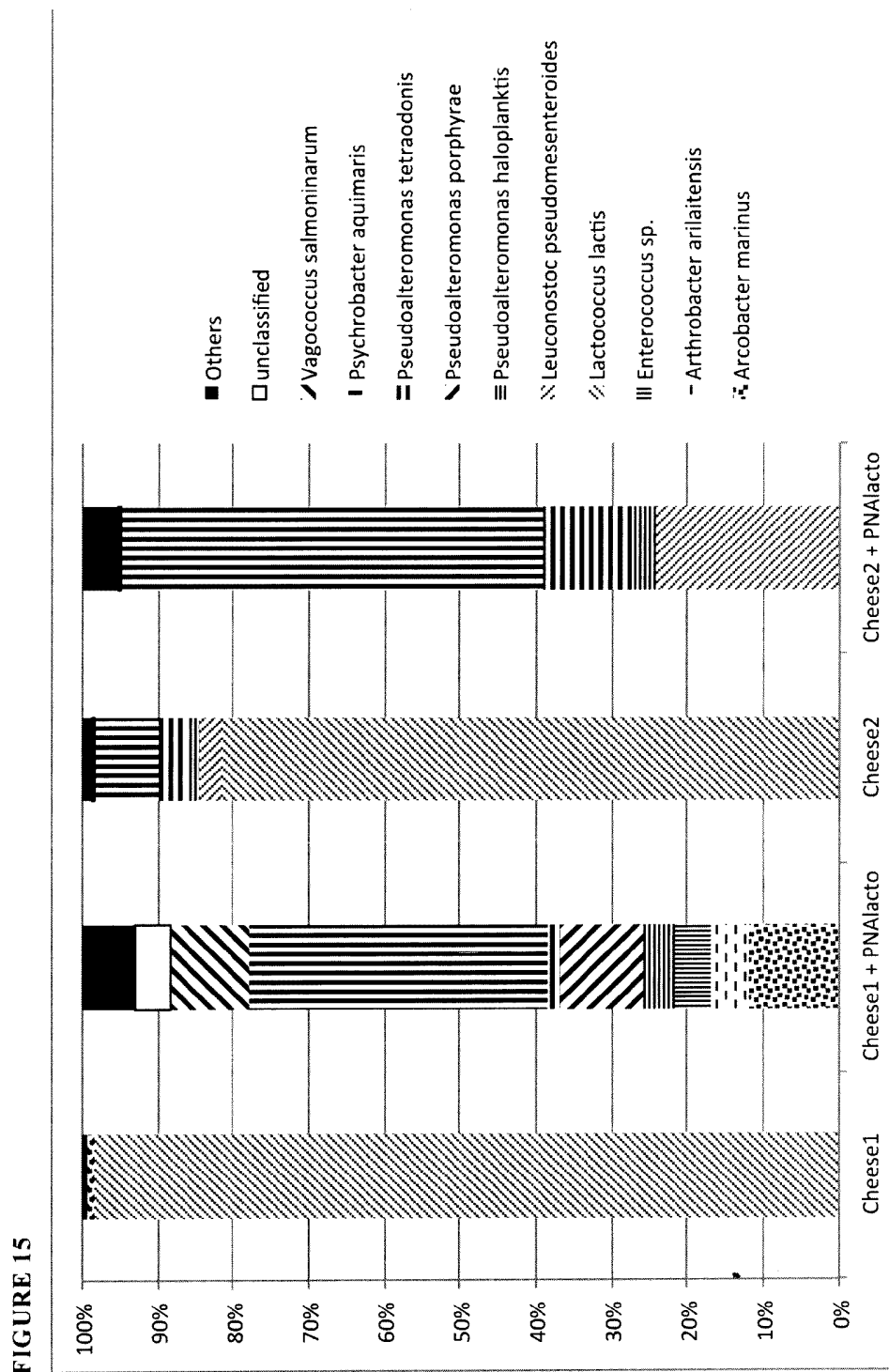


FIGURE 16

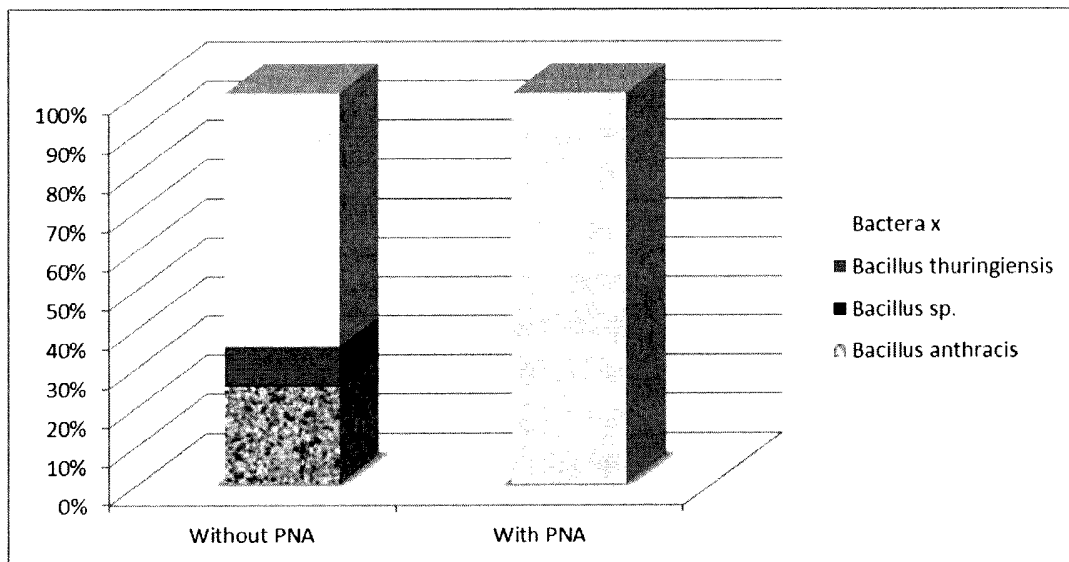


FIGURE 17

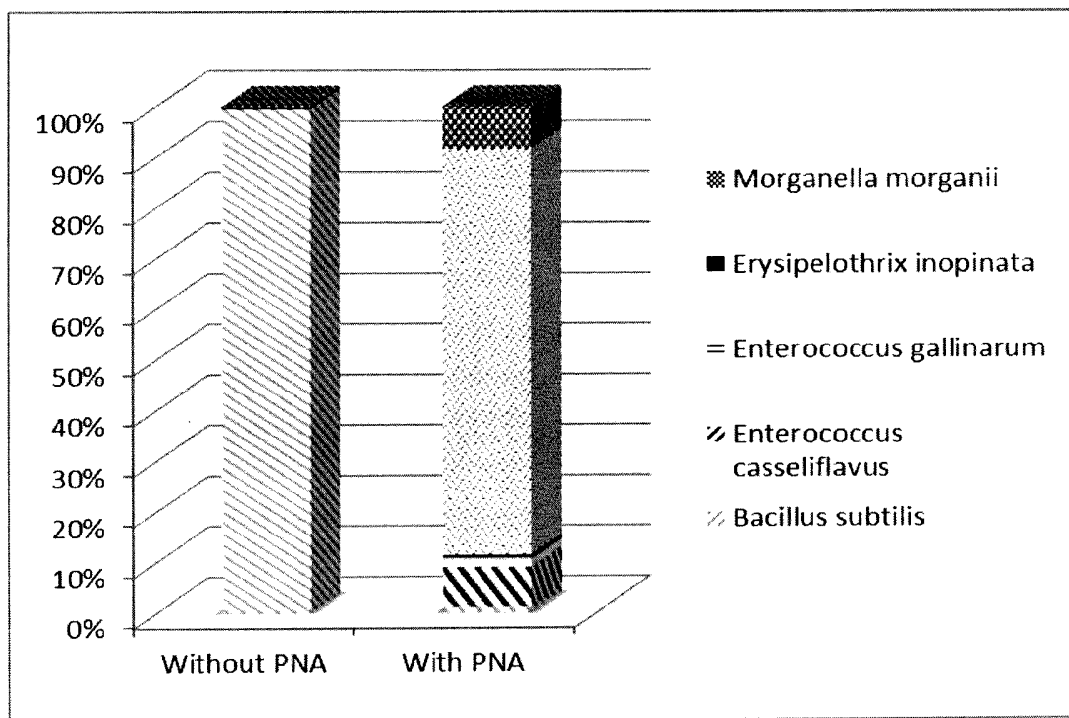


FIGURE 18

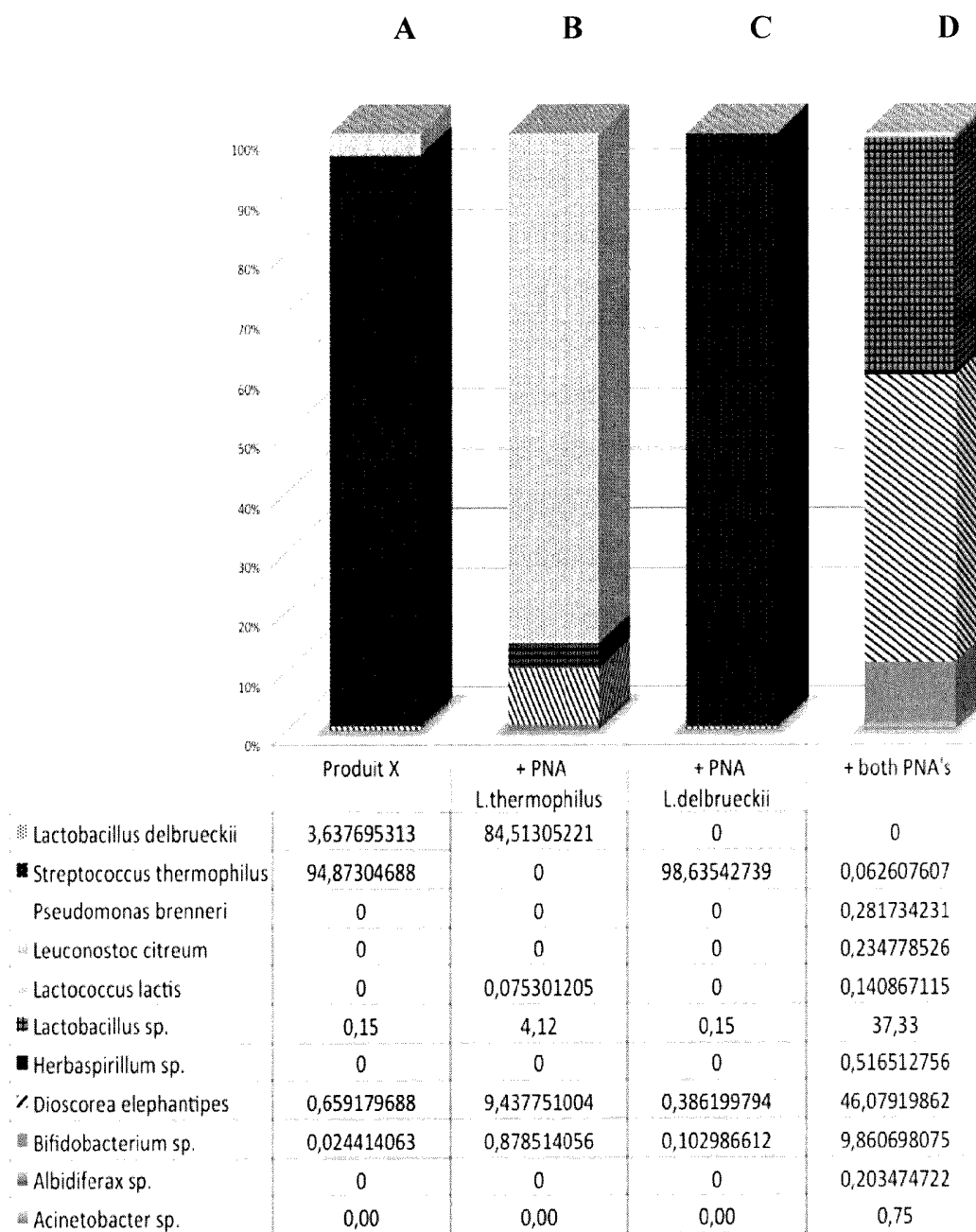


FIGURE 19

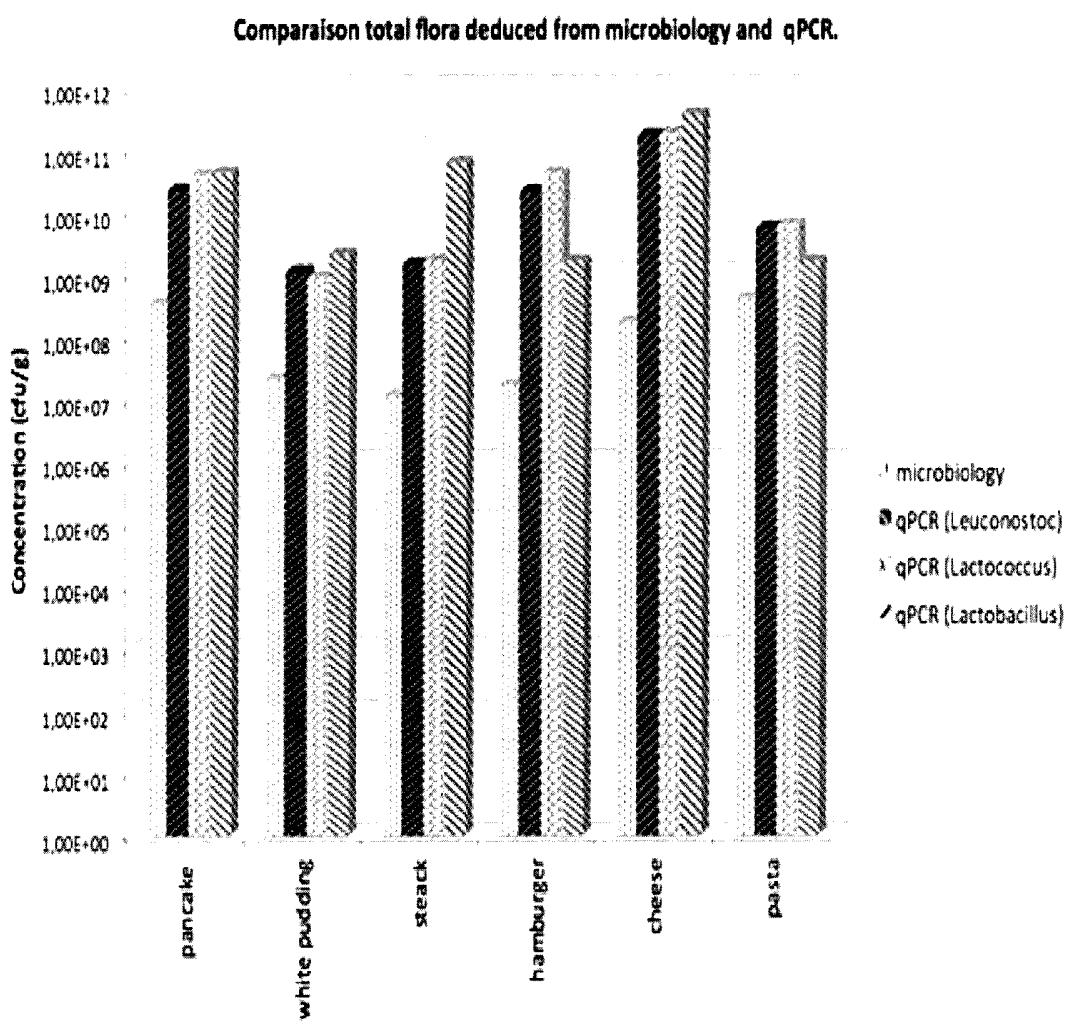


FIGURE 20 (A)

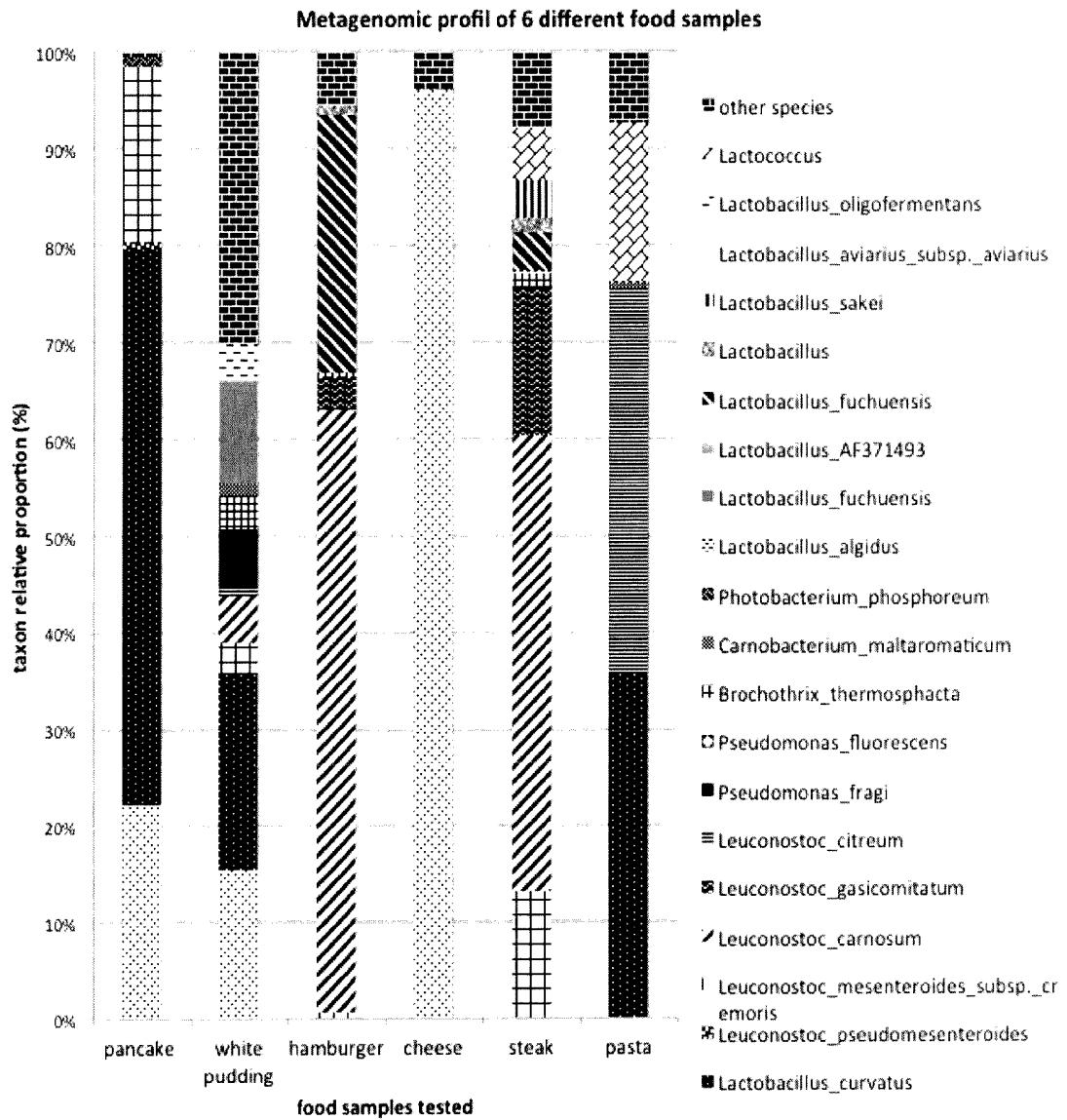
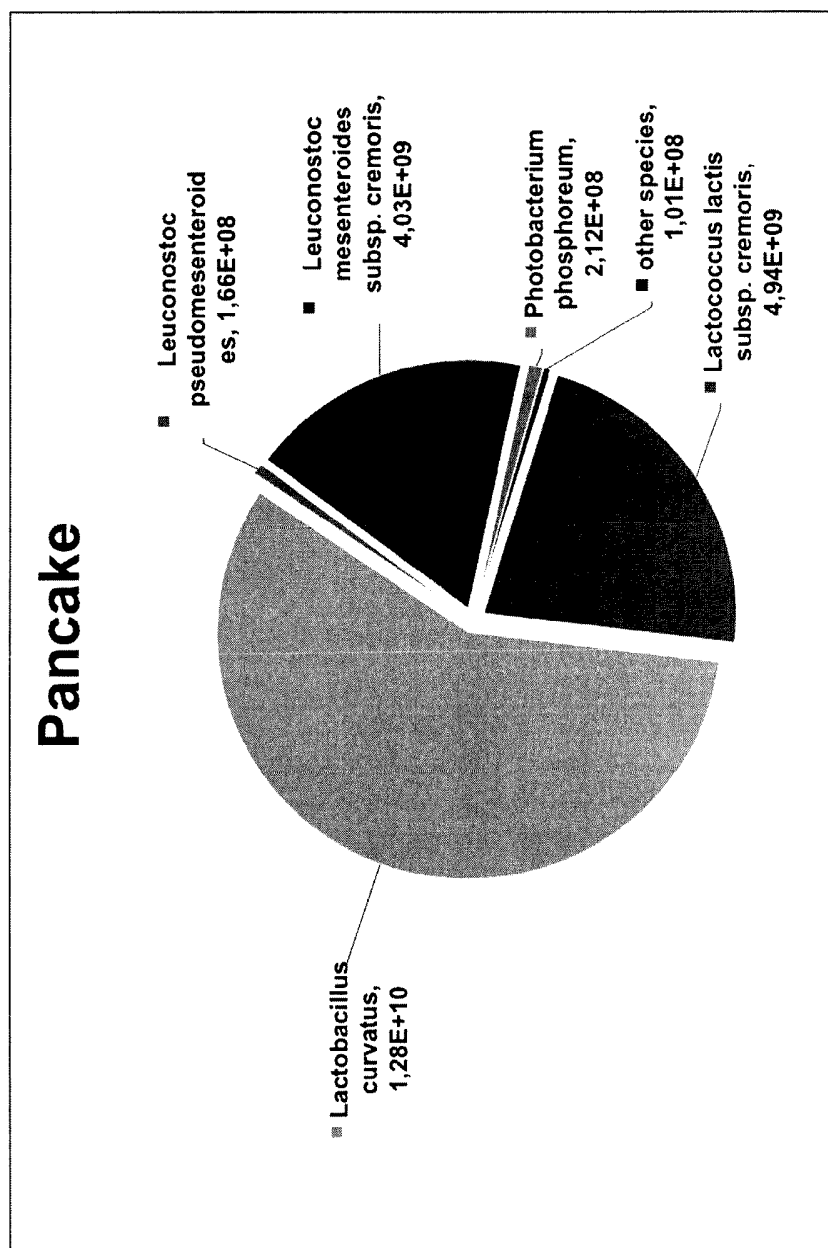


FIGURE 20 (B) (1/6)



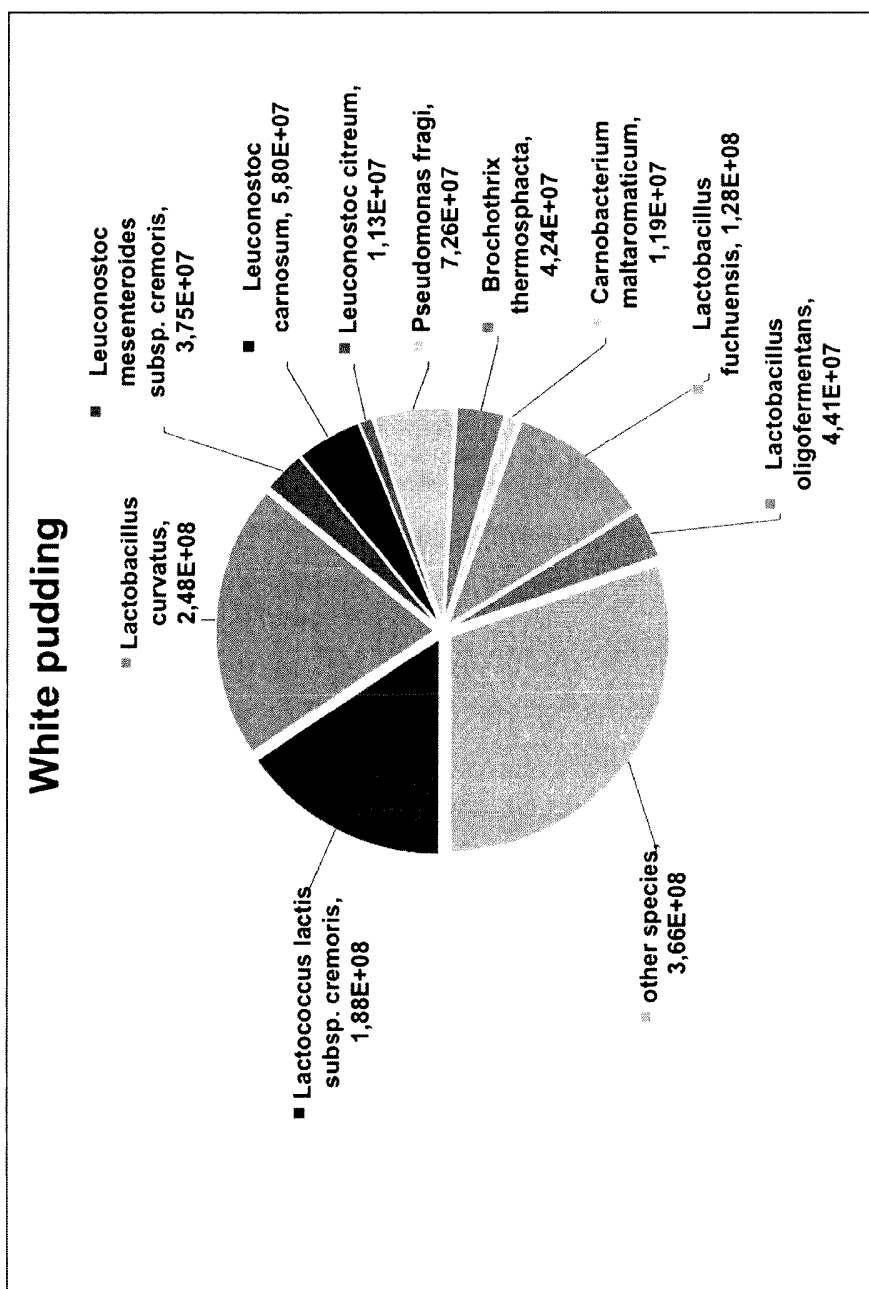


FIGURE 20 (B) (2/6)

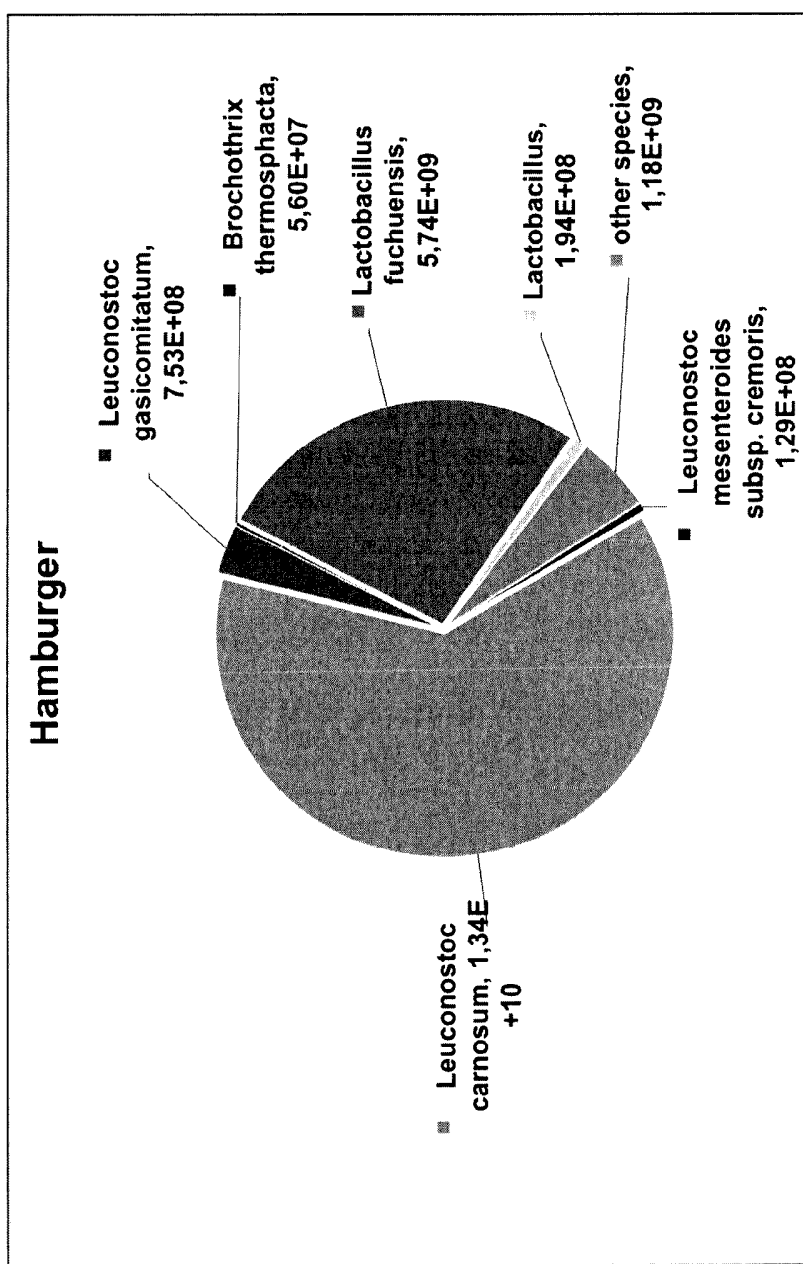


FIGURE 20 (B) (3/6)



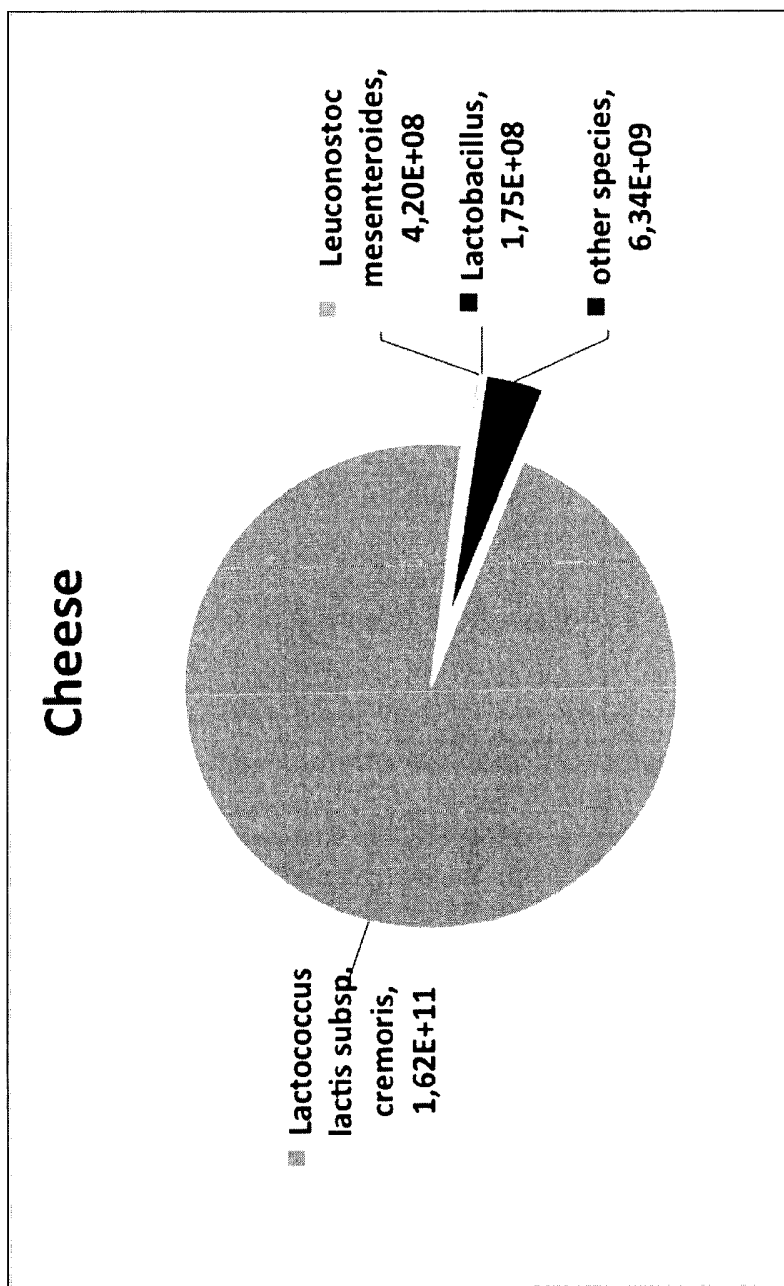
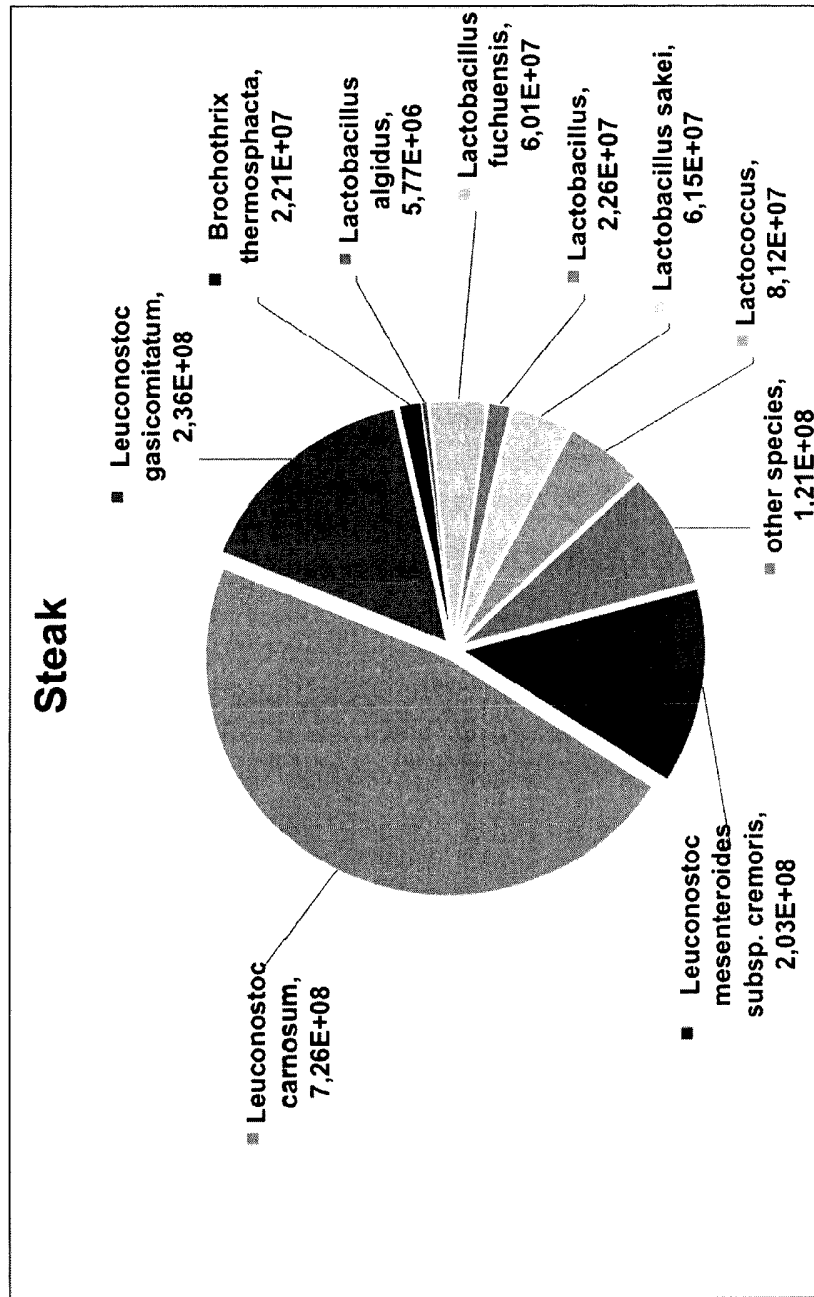


FIGURE 20 (B) (4/6)

FIGURE 20 (B) (5/6)



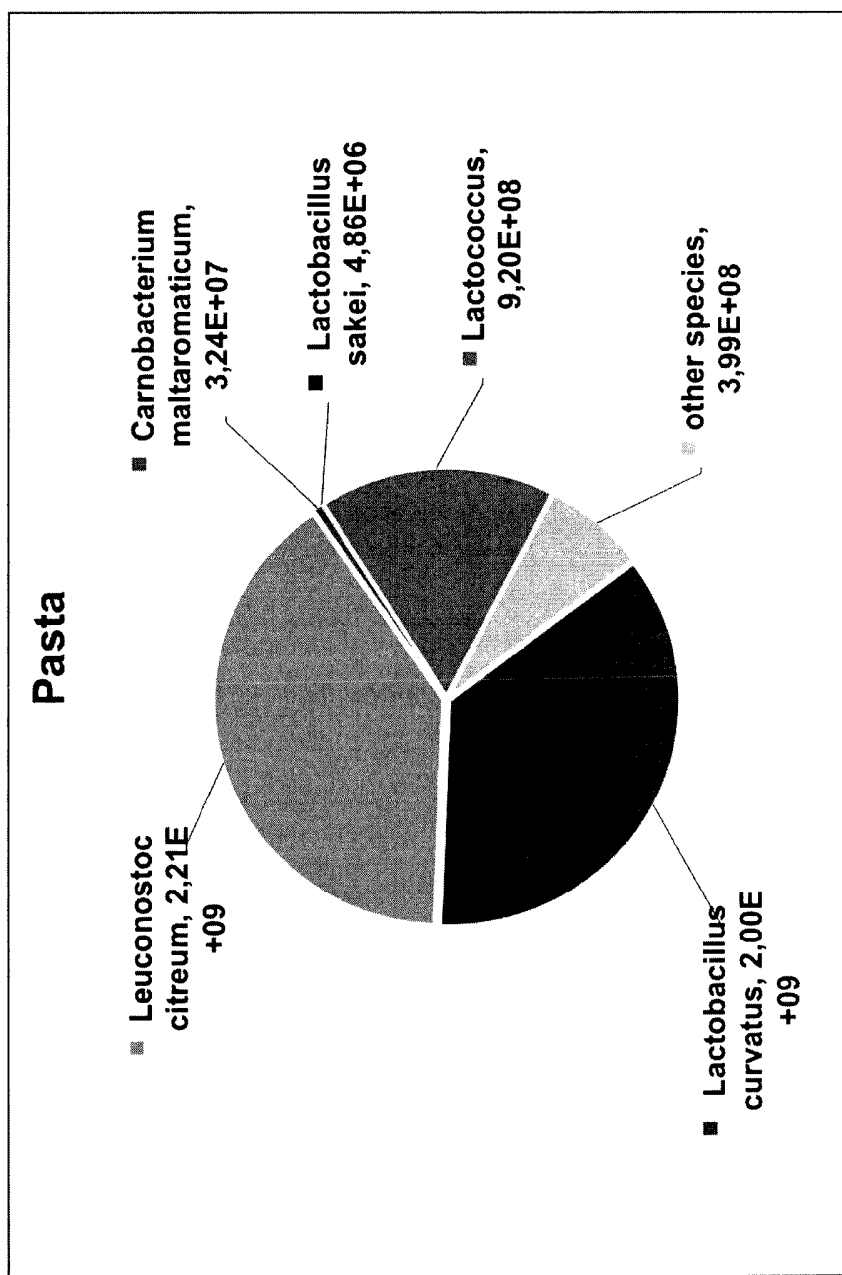


FIGURE 20 (B) (6/6)

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2014/077672

## Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:
  - a. (means)
    - on paper
    - in electronic form
  - b. (time)
    - in the international application as filed
    - together with the international application in electronic form
    - subsequently to this Authority for the purpose of search
2.  In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

# INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2014/077672

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> INV. C12Q1/68 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols) C12Q		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EUN-JIN PARK ET AL: "Bacterial community analysis during fermentation of ten representative kinds of kimchi with barcoded pyrosequencing", FOOD MICROBIOLOGY, ACADEMIC PRESS LTD, LONDON, GB, vol. 30, no. 1, 7 October 2011 (2011-10-07), pages 197-204, XP028445534, ISSN: 0740-0020, DOI: 10.1016/J.FM.2011.10.011 [retrieved on 2011-10-19] page 198 page 200 abstract  ----- -/--	1-30
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <span style="margin-left: 200px;"><input type="checkbox"/> See patent family annex.</span>		
* Special categories of cited documents :		
"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family	
Date of the actual completion of the international search  <p style="text-align: center;">12 March 2015</p>	Date of mailing of the international search report  <p style="text-align: center;">20/03/2015</p>	
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer  <p style="text-align: center;">Reuter, Uwe</p>	

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2014/077672

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>I. M. CARROLL ET AL: "Alterations in composition and diversity of the intestinal microbiota in patients with diarrhea-predominant irritable bowel syndrome", NEUROGASTROENTEROLOGY &amp; MOTILITY, vol. 24, no. 6, 20 February 2012 (2012-02-20), pages 521-e248, XP055119317, ISSN: 1350-1925, DOI: 10.1111/j.1365-2982.2012.01891.x page 522 - page 523; figure 1</p> <p style="text-align: center;">-----</p>	1-30
Y	<p>NOCKER ANDREAS ET AL: "Discrimination between live and dead cells in bacterial communities from environmental water samples analyzed by 454 pyrosequencing", INTERNATIONAL MICROBIOLOGY, SPRINGER VERLAG IBERICA, BARCELONA, ES, vol. 13, no. 2, 1 January 2010 (2010-01-01), pages 59-65, XP002671329, ISSN: 1139-6709, DOI: 10.2436/20.1501.01.111 page 60</p> <p style="text-align: center;">-----</p>	1-25,29
Y	<p>DEREK S LUNDBERG ET AL: "Practical innovations for high-throughput amplicon sequencing", NATURE METHODS, vol. 10, no. 10, 1 September 2013 (2013-09-01), pages 999-1002, XP055120581, ISSN: 1548-7091, DOI: 10.1038/nmeth.2634 page 999 page 1001 &amp; Derek S Lundberg ET AL: "Supplementary Information for: Practical innovations for high-throughput amplicon sequencing", Nature Methods, 1 September 2013 (2013-09-01), XP055143051, DOI: 10.1038/nmeth.2634 Retrieved from the Internet: URL: <a href="http://www.nature.com/nmeth/journal/v10/n10/extref/nmeth.2634-S1.pdf">http://www.nature.com/nmeth/journal/v10/n10/extref/nmeth.2634-S1.pdf</a> [retrieved on 2014-09-26]</p> <p style="text-align: center;">-----</p> <p style="text-align: center;">-/--</p>	1-30

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2014/077672

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>WINTZINGERODE VON F ET AL: "Peptide nucleic acid-mediated PCR clamping as a useful supplement in the determination of microbial diversity",  APPLIED AND ENVIRONMENTAL MICROBIOLOGY,  AMERICAN SOCIETY FOR MICROBIOLOGY, US,  vol. 66, no. 2,  1 February 2000 (2000-02-01), pages  549-557, XP002310906,  ISSN: 0099-2240, DOI:  10.1128/AEM.66.2.549-557.2000  page 555 - page 557</p> <p style="text-align: center;">-----</p>	1-30
A	<p>MASAO SAKAI ET AL: "Application of peptide nucleic acid (PNA)-PCR clamping technique to investigate the community structures of rhizobacteria associated with plant roots",  JOURNAL OF MICROBIOLOGICAL METHODS,  vol. 92, no. 3, 1 March 2013 (2013-03-01),  pages 281-288, XP055141983,  ISSN: 0167-7012, DOI:  10.1016/j.mimet.2012.09.036  page 284</p> <p style="text-align: center;">-----</p>	1-30