Clinical detection of human probiotics and human pathogenic bacteria by using a novel high-throughput platform based on next generation sequencing

Chih-Min Chiu[1], Feng-Mao Lin[1, §], Tzu-Hao Chang[2, §], Wei-Chih Huang[1], Chao Liang[1], Wei-Yun Wu[1], Tzu-Ling Yang[1], Shun-Long Weng[1,4,5,6*], Hsien-Da Huang[1,3*]

[1]Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsin-Chu 300, Taiwan
[2]Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei, Taiwan
[3]Department of Biological Science and Technology, National Chiao Tung University, Hsin-Chu 300, Taiwan
[4]Department of Obstetrics and Gynecology, Hsinchu Mackay Memorial Hospital, Hsinchu, Taiwan
[5]Mackay Medicine, Nursing and Management College, Taipei, Taiwan
[6]Department of Medicine, Mackay Medical College, New Taipei City, Taiwan

[§]These authors contributed equally to this work
*Corresponding authors

Email addresses:
CHC: citrate34@gmail.com
FML: kiralintw@gmail.com
THC: kevinchang@tmu.edu.tw
WCH: gohk0958@msn.com
CL: gsuper3150631@gmail.com
WYW: ukpiyo@gmail.com
TLY: tzulingyang@gmail.com
SLW: a4467@ms7.mmh.org.tw
HDH: bryan@mail.nctu.edu.tw

Abstract. The human body plays host to a vast array of bacteria which are harmful or beneficial. Next generation sequencing technology have increased its accuracy in identifying bacteria. This work develops a novel platform for rapidly detecting probiotics and pathogens based on sequencing results of 16S rRNA. A database that recorded the species of probiotics and pathogens from literature was constructed, along with a modified Smith-Waterman algorithm for assigning the taxonomy of the sequenced 16S rRNA sequences. A bacteria disease risk model for seven diseases was constructed based on 98 samples. Applicability of the proposed platform is demonstrated by collecting the microbiome in human gut of 13 samples. The proposed platform provides a relatively easy means of identifying a certain amount of bacteria and their species for clinical microbiology applications. Detecting how probiotics and pathogens inhabit humans and affect their health significantly contributes to develop a diagnosis and treatment method.

1 Background

High throughput sequencing can analyze a large amount of sequences, enabling sequencing of 16S rRNA to identify complex bacteria species of pathogens and probiotic bacteria. Many naturally occurring bacteria form a complex population in the environment. The human body plays host to a vast array of bacteria, found in oral cavities, skin, gastrointestinal tract and the vagina. Some bacteria are harmful while others are beneficial to the host.

As a pathogen, a microorganism causes disease in its host. Among the many examples of bacterial pathogen include *Corynebacterium diphtheria* (causes diphtheria), *Listeria monocytogenes* (causes food poisons), and *Legionella pneumophila* (causes Legionnaires' disease).

As microorganisms, probiotics benefit the host. Probiotics has received considerable attention in recent years. A FAO report in 2001 [1] cited the advantages of probiotics as increasing immunity [2], reducing gastrointestinal discomfort [3] and protecting the flora within urogenital tract [4]. As is well known, probiotics can ameliorate symptoms of diseases [5] and reduce the risk of suffering from diseases [6, 7].

Despite the availability of many approaches to identify probiotic and pathogens, most of them are only applicable to specific and cultivable bacteria and are also time consuming. For instance, conventional methods detect growth of cultured bacteria in approximately two days, or an additional five days to obtain no-growth culture results [8], which is laborious. Additionally, some bacteria cannot be cultured [9], subsequently increasing the difficulty of specifying pathogenic bacteria. Moreover, it is hard to determine whether an infection is caused by one or more bacteria types.

Capable of identifying bacteria on a molecular level, 16S rRNA sequences can also detect uncultivable bacteria [10]. Use of 16S rRNA sequencing can overcome some problems of conventional culture method [11]. Although a more effective means of identifying bacteria than conventional culture method, 16S rRNA sequencing takes a considerable amount of time in amplifying DNA sequences [12]. Sanger sequencing, also known as “first-generation”, or “conventional” sequencing, is used for DNA sequencing for almost two decades. Compared to next generation sequencing (NGS), NGS is more efficient than the Sanger method, owing to its ability to analyze large-scale sequences quicker, enable massively parallel analysis, reduce reagent costs, reduce the size of sample components and perform high throughput [13]. Additionally, NGS is more appropriate for bacteria community identification because of its ability to generate millions of reads per sample than Sanger method which generates one read per sample. Also, NGS of 16S rRNA can more easily identify cultivable and uncultivable bacteria [10].

Recent advances in sequencing technology and Bioinformatics approaches have increased its accuracy in distinguishing bacteria. Based on high throughput sequencing technology, this work identifies 16S rRNA sequences of bacteria and analyzes bacteria species. High-throughput sequencing can sequence a large number of 16S rRNA sequences more efficiently and allow researchers to acquire information in order to identify pathogens and probiotic bacteria.

2 Methods

Figure 1 illustrates the bioinformatics system flow of the proposed platform including analysis pipeline of NGS, construction of probiotics and pathogens database, bacterial disease risk model evaluation and the application of individualized bacteria sequencing profile. The detailed components in the proposed platform are described below.

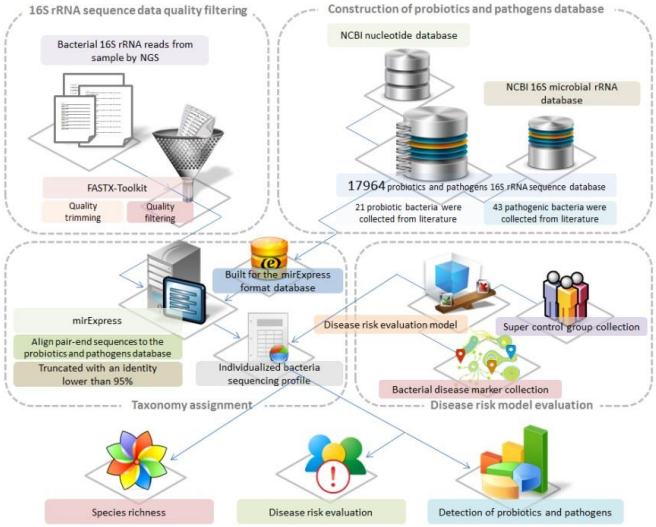

Fig. 1. System flow of bioinformatics analysis in the proposed platform

2.1 Sample collection, preparation and sequencing for V4 region of 16 S ribosomal DNA

In this study, stool samples of 98 Taiwan volunteers were gathered. The samples were collected by Sigma-transwab (Medical Wire) into a tube with Liquid Amies Transport Medium, and stored at 4°C until processing. DNA was extracted directly on stool samples by using a QIAamp DNA Stool Mini Kit (Qiagen). The PCR primers, F515 (5'-GTGCCAGCMGCCGCGGTAA-3') and R806 (5'-GGACTACHVGGGTWTCTAAT-3'), were designed to amplify the V4 domain of bacterial 16S ribosomal DNA as described previously [14]. After PCR amplification was performed, Illumina adapters were attached to the amplicons using the Illumina TruSeq DNA Sample Preparation v2 Kit for V4 library preparation. Purified libraries were applied for cluster generation and sequencing on the MiSeq system.

2.2 16S rRNA (rDNA) sequence data quality filtering

The raw fastq files obtained by Illumina sequencing machine were quality-filtered using the FASTX-Toolkit[1]. The paired-end 150bp reads were performed using the minimum acceptable phred quality score of 20, as well as the 70% of bases that must exceed 20 phred quality score. Sequence shorter than 100 nucleotides would be omitted after quality trimming from reads tail. Notably, reads containing ambiguous characters were discarded.

2.3 Construct a curated 16S rRNA database of probiotics and pathogens and Taxonomy mapping.

The species of probiotics were collected from both literature[5, 7] and the claims of official departments, such as Health Canada [15] and Taiwan Food and Drug Administration [16]. The species of pathogens were collected from literature [17-28]. A total of 21 probiotics and 99 bacterial pathogens were collected. The 16S rRNA sequences of probiotics and pathogens used for taxonomy mapping were retrieved from the NCBI nucleotide database, NCBI 16S microbial rRNA database, Greengenes database[29] and SILVA[30]. Following sequence data collection, we assemble partial sequences which used the same species classification and removed redundant sequences. Additionally, we also removed the unique sequence from only one research support with 3% similarity which shared the same species classification with other sequence.

To generate taxonomy assignments, the proposed platform invoked a modified Smith-Waterman algorithm from miRExpress [31], which can compare pairs of sequences in parallel, for mapping reads to taxons. miRExpress was designed for identifying the best similarity between sequencing read and reference sequence. In our model, it was modified for identifying multiple hits of 16S rRNA sequence mapping results with similarity threshold 0.95. The SAM format [32] was used to replace the original miRExpress output format for storing alignment results. miRExpress was originally designed for dealing with single-end sequencing data. Therefore, the additional program was added for processing paired-end sequencing data. The probiotics and pathogens 16S rRNA sequence from our database were built in FASTA format. Following quality filtering, all paired-end sequences were aligned to the probiotics and pathogens database with whole read aligned from one end to the other end. Reads were then truncated with an identity lower than 95%, according to previous research in order to achieve a better compromise between sequences from PCR sequencing errors and taxonomic relatedness [33].

2.4 Bacterial disease risk evaluation model construction

For studying the associations between bacteria and diseases, related information were collected from literatures. The associations between bacteria and seven diseases, constipation [34-36], obesity [37-41], irritable bowel syndrome (IBS) [36, 42-47],

[1] http://hannonlab.cshl.edu/fastx_toolkit/index.html

ulcerative colitis (UC) [42, 48-50], colon cancer (CC) [51-53], Atopic Dermatitis (AD) and Allergic rhinitis (AR), were collected (positive correlation and negative correlation data) and the individual risk of disease was evaluated. The association data were majorly collected from case-control studies which the quantities of bacteria were obtained from NGS data, and few well-known bacteria validated by multiple studies through cultural experiments were also included. We further eliminated some conflicted data with both positive and negative correlation between bacteria and disease in different studies.

Health Asians stool samples of 98 Taiwan volunteers were gathered. Following deep sequencing and sequencing data processing, the proportion of 78 bacteria from control group was applied as risk markers (constipation: 6, obesity: 9, IBS: 17, UC: 10, CC: 28, AD: 4, AR: 4) to predict disease risk to seven diseases in this study. The disease risks were evaluated from markers and disease through binomial test. Let N risk markers be the number of trials in each disease, and let X be the counted number of successfully exceeding risk boundary (high risk) of each bacterial risk marker. Therefore, X/N (hypothesized probability) represented the numbers of high risk bacteria were detected in a random subject, or regarded as a binomial random variable. Hypothesized probabilities of seven diseases (0.05051 for constipation, 0.07239 for obesity, 0.06952 for IBS, 0.05227 for UC, 0.09280 for CC, 0.04924 for AD, 0.05114 for AR) were generated by the numbers of bacterial risk markers which successfully exceeded risk boundary in each disease of 98 samples control group.

Fence of boxplot was used to generalize a risk boundary of a specific marker from distribution of proportional data in 98 samples control group. If a bacterial risk marker recorded as positive correlation, the upper fence of proportion data in control group were defined as risk boundary. Personal proportion data exceeded this boundary were defined as successfully exceeding. The lower fence was used when a marker negatively correlated with a disease. If personal proportion data was lower than this boundary, the result was defined as a success.

3 Results

3.1 Platform application: gut probiotics and pathogens detection

To demonstrate the capability of the proposed platform, 13 human fecal samples were collected individually from participants for sequencing on MiSeq system. Sequencing generated 1,597,816 paired-end sequencing reads of the V4 region of the 16S rRNA that passed the quality filtering. An average of 122,908 reads was acquired per sample. 151,629 reads from 13 samples were then assigned to probiotics and pathogens by using mirExpress.

Figure 2A illustrated the percentage of probiotics detected by the proposed platform. The top three probiotics identified in the 13 samples are *Lactococcus salivarius*, *Streptococcus thermophilus* and *Bifidobacterium longum*. Figure 2B was depicted the proportion of pathogens detected by the proposed platform. The top three pathogens are *Escherichia coli, Salmonella enterica* and *Haemophilus influenza*.

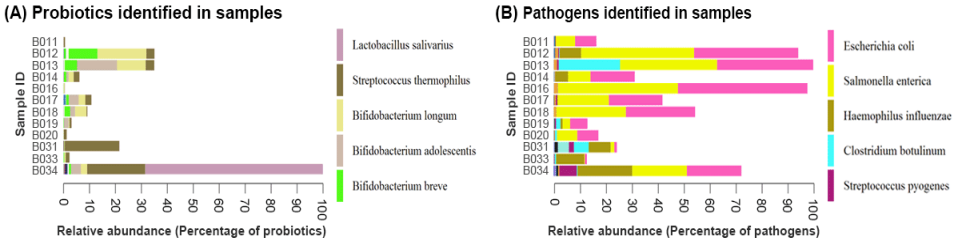

Fig. 2. The percentage of probiotics and pathogens were identified in the samples in the case study.

The results of disease risk evaluations indicated that only two of 13 samples had similar distributions of three diseases with 98 samples control group. Sample B031 had reached the significance level with P-value 0.0333 and 0.0121< 0.05 of distribution in irritable bowel syndrome and colorectal cancer respectively compared to 98 samples control group using binomial test. Sample B034 had reached the significance level with P-value 0.0025 and 0.0121< 0.05 of distribution in obesity and colorectal cancer. Evaluating by the association of bacterial risk markers and disease, the results suggested that these two samples had higher risk than 98 samples control group in irritable bowel syndrome, colorectal cancer and obesity. Their enterotypes of gut probiotics and pathogens may be one of risk factors which would cause disease.

3.2 Reproducibility and accuracy evaluation

Reproducibility of NGS was evaluated by selecting four samples to perform 2-replicate experiments. For evaluate 2-replicate reproducibility correctly, the phylogenetic relationship must take into account. UniFrac [54] was used for calculating a distance measure between organismal communities (pairs of samples) using phylogenetic information. Results of each sample pair (replicate 1 and 2) closely resemble each other. The similarity of UniFrac distance of each sample pair is higher than 0.96 (0.9617 for B014, 0.9872 for B018, 0.9914 for B020, 0.9722 for B033), implying that the analysis results are reproducible.

Next, accuracy of NGS is evaluated by adding *Lactobacillus reuteri* to a stool sample (B050). Sample B050 contains 24,408 assigned taxons, and *Lactobacillus reuteri* has no detected count. Whether the counts of this species in positive control sample (B050S_L) are elevated must be determined. Analysis results indicate that 27,113 taxons are detected in sample B050S_L. In fact, the detected counts of *Lactobacillus reuteri* in sample B050S_L are 1,430, and the percentage of *Lactobacillus reuteri* markedly increases from 0% to 5%.

In summing up the above results, NGS is accurate and reproducible in terms of detecting the quantities of bacterial species of the proposed platform. The results evaluate the accuracy and feasibility of NGS techniques in order to identify probiotics and pathogens. While requiring only about one day for detection, not limited in identifying certain bacteria, the proposed platform can detect and quantify multiple bacteria simultaneously.

4 Discussion

Owing to the constraint of costs and technical limitations, 16S rRNA sequences obtained in most databases are partial sequences. Many studies thus assign taxonomy by using partial 16S rRNA sequences. In our probiotics and pathogens 16S rRNA sequence database, 17,964 sequences are collected from NCBI nucleotide database, NCBI 16S microbial rRNA database, Greengenes database and SILVA. Our probiotics and pathogens 16S rRNA database contains less than 39% of 16S rRNA sequences which are longer than 1400 bps. Only 9% of the sequences are close to full length.

This work extracts the V4 region from full length 16S rRNA of microbiome in the human gut as a platform application. Some 16S rRNA variable regions are more dependable than other regions in assigning taxonomy like V3 and V4 [55, 56] ; in addition, some 16S rRNA variable regions are much conserved. The proportion and diversity of probiotics and pathogens may be made diverse by using different 16S rRNA variable regions. The proposed platform is also applicable to other 16S rRNA variable regions for taxonomy assignment. Importantly, a more appropriate region than others must be selected to produce an outcome that is close to full length 16S rRNA sequence.

This work further attempts to collect common probiotics and pathogens from the literature, although it may be incomplete. Nevertheless, recent advances in sequencing technology make it possible to identify and define an increasing number of bacteria, implying an obvious increase in the number of identified probiotics and pathogens in the future. Efforts are underway in our laboratory to update the list of used probiotics and pathogens.

Previous studies [57-59] identified pathogen or probiotic bacteria by using antibody, 16S rRNA gene microarrays, fluorescence *in situ* hybridization (FISH) and proteomic methods. In this work, the proposed platform can detect various pathogens and probiotics based on 16S rRNA (rDNA) sequences of bacteria using NGS and Bioinformatics method. An average of 122,908 reads was acquired per sample in this work. It is doubt that the sequencing depth is enough to detect a small amount of probiotics and pathogens. Although increasing the coverage of sequencing can advance the sensitivity of detecting probiotics and pathogens, the sequencing cost will increase. It is important to work out an appropriate coverage of sequencing for detecting probiotics and pathogens.

The results of disease risk evaluations revealed that most of 13 samples did not have resembled distributions of bacteria markers with control group. Only two samples had reached the significance level of distributions. The reason for the phenomenon may be the overlapped bacteria markers between diseases. There are 28 markers used in colorectal cancer, and 17 markers used in irritable bowel syndrome. Six markers are overlapped. For sample B031, the significant distributions in colorectal cancer were partly contributed to the significance in irritable bowel syndrome owing to the overlapped markers. Similarly, two overlapped markers for sample B034 were in colorectal cancer and obesity. In this kind of speculation, the influence of colorectal cancer to irritable bowel syndrome would be six (overlapped markers of CC and IBS) over seventeen (markers of IBS), and the influence of colorectal cancer to obesity would be two (overlapped markers of CC and obesity)

over nine (markers of obesity). In addition, the influence of colorectal cancer to constipation and ulcerative colitis would be one over six and two over ten, respectively.

In addition to that some bacteria markers in species level are belong to the marker of genus level and species level, genus marker and species markers may have some associations that affecting the distributions mutually. Continually, collecting more markers and evaluating the distributions with markers in the same level are required for constructing a global prediction model in Taiwanese.

5 Conclusions

Probiotics and pathogens in human oral, nasal, skin, gut and urogenital affect human health. Also, the quality of health is based on the amount of probiotics and pathogens in the human body. Detecting how probiotics and pathogens inhabit humans and affect their health significantly contributes to develop a diagnosis and treatment method. The proposed platform can reduce the time of traditional bacteria culture method by NGS, as well as detect rapidly the proportion of probiotics and pathogens (including uncultivable pathogens) in the human body by using probiotics and pathogens 16S rRNA database.

In clinical application, the proposed platform provides further insight into the cause of disease as well as the relation of probiotics, pathogens and disease by rapid detection. For infectious patients, once the bacterial species of pathogens and quantity of each species are identified, the type and administration of antibiotics can be adjusted. In health care, the proposed platform allows researchers to determine whether the intake of probiotics impacts the human body. The number of probiotics in the body can serve as an index to determine whether probiotics intake can improve human health. Furthermore, changes in the amount of probiotics administered can serve as a health-evaluating index [35, 38, 44, 47, 48].

This work constructed a bacterial disease risk evaluation model for seven diseases and develops a novel platform to identify probiotics and pathogens in the human samples by using NGS and Bioinformatics approach, making it highly promising for clinical microbiology applications as well as basis of clinical diagnosis. To construct a global prediction model in human, this preliminary study will be continuously extended for more bacterial disease markers. For more comprehensive applications, this work will also continuously collect more Taiwan health samples of other parts of human body as control group data by using NGS to analyze bacteria composition.

Competing interests

The authors declare no competing interests.

Authors' contributions

HDH conceived and supervised the study. CMC were responsible for the design, computational analyses, implementation of the system, and drafting the manuscript. FML, THC, CL, TLY, SLW, WCH and WYW were in charge of manuscript revision and data update. All authors read and approved the final manuscript**.**

Acknowledgements

The authors would like to thank the National Science Council of the Republic of China for financially supporting this research under Contract No. NSC 101-2311-B-009-003-MY3 and NSC 100-2627-B-009-002. This work was supported in part by the UST-UCSD International Center of Excellence in Advanced Bio-engineering sponsored by the Taiwan National Science Council I-RiCE Program under Grant Number: NSC 101-2911-I-009-101, and Veterans General Hospitals and University System of Taiwan (VGHUST) Joint Research Program under Grant Number: VGHUST101-G5-1-1. This work was also partially supported by MOE ATU.

References

1. FAO/WHO: **Health and nutritional properties of probiotics in food including powder milk with live lactic acid bacteria**. In: *Report of a joint FAO/WHO expert consultation on evaluation of health and nutritional properties of probiotics in food including powder milk with live lactic acid bacteria.* http://www.who.int/foodsafety/publications/fs_management/en/probiotics.pdf; 2001.
2. Isolauri E, Sutas Y, Kankaanpaa P, Arvilommi H, Salminen S: **Probiotics: effects on immunity**. *The American journal of clinical nutrition* 2001, **73**(2 Suppl):444S-450S.
3. Gopal PK, Prasad J, Smart J, Gill HS: **In vitro adherence properties of Lactobacillus rhamnosus DR20 and Bifidobacterium lactis DR10 strains and their antagonistic activity against an enterotoxigenic Escherichia coli**. *International journal of food microbiology* 2001, **67**(3):207-216.
4. Reid G: **Probiotic agents to protect the urogenital tract against infection**. *The American journal of clinical nutrition* 2001, **73**(2 Suppl):437S-443S.
5. Wan YM, Zhu YQ, Xia B, Luo J: **Treating TNBS-induced colitis in rats with probiotics**. *The Turkish journal of gastroenterology : the official journal of Turkish Society of Gastroenterology* 2011, **22**(5):486-493.
6. Taipale T, Pienihakkinen K, Isolauri E, Larsen C, Brockmann E, Alanen P, Jokela J, Soderling E: **Bifidobacterium animalis subsp. lactis BB-12 in reducing the risk of infections in infancy**. *The British journal of nutrition* 2011, **105**(3):409-416.
7. Wickens K, Black PN, Stanley TV, Mitchell E, Fitzharris P, Tannock GW, Purdie G, Crane J: **A differential effect of 2 probiotics in the prevention of eczema and atopy: a double-blind, randomized, placebo-controlled trial**. *The Journal of allergy and clinical immunology* 2008, **122**(4):788-794.
8. Jordan JA, Durso MB: **Real-time polymerase chain reaction for detecting bacterial DNA directly from blood of neonates being evaluated for sepsis**. *The Journal of molecular diagnostics : JMD* 2005, **7**(5):575-581.

9. Wilson M: **Bacteriology of Humans: An Ecological Perspective**: Wiley-Blackwell; 2008.
10. Goldenberger D, Kunzli A, Vogt P, Zbinden R, Altwegg M: **Molecular diagnosis of bacterial endocarditis by broad-range PCR amplification and direct sequencing**. *J Clin Microbiol* 1997, **35**(11):2733-2739.
11. Woo PC, Lau SK, Woo GK, Fung AM, Ngan AH, Hui WT, Yuen KY: **Seronegative bacteremic melioidosis caused by Burkholderia pseudomallei with ambiguous biochemical profile: clinical importance of accurate identification by 16S rRNA gene and groEL gene sequencing**. *J Clin Microbiol* 2003, **41**(8):3973-3977.
12. Fox GE, Magrum LJ, Balch WE, Wolfe RS, Woese CR: **Classification of methanogenic bacteria by 16S ribosomal RNA characterization**. *Proc Natl Acad Sci U S A* 1977, **74**(10):4537-4541.
13. Metzker ML: **Sequencing technologies - the next generation**. *Nat Rev Genet* 2010, **11**(1):31-46.
14. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R: **Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample**. *Proc Natl Acad Sci U S A* 2011, **108 Suppl 1**:4516-4522.
15. Lin J: **Effect of Antibiotic Growth Promoters on Intestinal Microbiota in Food Animals: A Novel Model for Studying the Relationship between Gut Microbiota and Human Obesity?** *Frontiers in microbiology* 2011, **2**:53.
16. Khamis A, Raoult D, La Scola B: **Comparison between rpoB and 16S rRNA gene sequencing for molecular identification of 168 clinical isolates of Corynebacterium**. *Journal of clinical microbiology* 2005, **43**(4):1934-1936.
17. Kenneth Todar PD: **Bordetella pertussis and Whooping Cough**. In: *Online Textbook of Bacteriology*. 2008.
18. Seleem MN, Boyle SM, Sriranganathan N: **Brucella: a pathogen without classic virulence genes**. *Veterinary microbiology* 2008, **129**(1-2):1-14.
19. Zhan P, Suo LJ, Qian Q, Shen XK, Qiu LX, Yu LK, Song Y: **Chlamydia pneumoniae infection and lung cancer risk: a meta-analysis**. *Eur J Cancer* 2011, **47**(5):742-747.
20. Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, Basham D, Chillingworth T, Davies RM, Feltwell T, Holroyd S *et al*: **The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences**. *Nature* 2000, **403**(6770):665-668.
21. Guidoboni M, Ferreri AJ, Ponzoni M, Doglioni C, Dolcetti R: **Infectious agents in mucosa-associated lymphoid tissue-type lymphomas: pathogenic role and therapeutic perspectives**. *Clinical lymphoma & myeloma* 2006, **6**(4):289-300.
22. Richard A. Harvey Ph.D. PCC, Bruce D. Fisher: **Lippincott's Illustrated Reviews: Microbiology**, Second edn; 2007.
23. Fredlund H, Falk L, Jurstrand M, Unemo M: **Molecular genetic methods for diagnosis and characterisation of Chlamydia trachomatis and Neisseria gonorrhoeae: impact on epidemiological surveillance and interventions**. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica* 2004, **112**(11-12):771-784.
24. Maley MW, Kociuba K, Chan RC: **Prevention of laboratory-acquired brucellosis: significant side effects of prophylaxis**. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2006, **42**(3):433-434.
25. Berdichevski T, Barshack I, Bar-Meir S, Ben-Horin S: **Pseudomembranes in a patient with flare-up of inflammatory bowel disease (IBD): is it only Clostridium difficile or is it still an IBD exacerbation?** *Endoscopy* 2010, **42 Suppl 2**:E131.
26. Satterfield BA, Stewart AF, Lew CS, Pickett DO, Cohen MN, Moore EA, Luedtke PF, O'Neill KL, Robison RA: **A quadruplex real-time PCR assay for rapid detection and differentiation of the Clostridium botulinum toxin genes A, B, E and F**. *Journal of medical microbiology* 2010, **59**(Pt 1):55-64.
27. Mantur BG, Amarnath SK, Shinde RS: **Review of clinical and laboratory features of human brucellosis**. *Indian journal of medical microbiology* 2007, **25**(3):188-202.

28. Brook I: **The role of anaerobic bacteria in cutaneous and soft tissue abscesses and infected cysts**. *Anaerobe* 2007, **13**(5-6):171-177.
29. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL: **Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB**. *Appl Environ Microbiol* 2006, **72**(7):5069-5072.
30. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO: **The SILVA ribosomal RNA gene database project: improved data processing and web-based tools**. *Nucleic acids research* 2013, **41**(D1):D590-596.
31. Wang WC, Lin FM, Chang WC, Lin KY, Huang HD, Lin NS: **miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression**. *BMC bioinformatics* 2009, **10**:328.
32. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools**. *Bioinformatics* 2009, **25**(16):2078-2079.
33. Hummelen R, Fernandes AD, Macklaim JM, Dickson RJ, Changalucha J, Gloor GB, Reid G: **Deep sequencing of the vaginal microbiota of women with HIV**. *PloS one* 2010, **5**(8):e12078.
34. Chmielewska A, Szajewska H: **Systematic review of randomised controlled trials: probiotics for functional constipation**. *World J Gastroenterol* 2010, **16**(1):69-75.
35. Chassard C, Dapoigny M, Scott KP, Crouzet L, Del'homme C, Marquet P, Martin JC, Pickering G, Ardid D, Eschalier A *et al*: **Functional dysbiosis within the gut microbiota of patients with constipated-irritable bowel syndrome**. *Aliment Pharmacol Ther* 2012, **35**(7):828-838.
36. Lyra A, Rinttila T, Nikkila J, Krogius-Kurikka L, Kajander K, Malinen E, Matto J, Makela L, Palva A: **Diarrhoea-predominant irritable bowel syndrome distinguishable by 16S rRNA gene phylotype quantification**. *World J Gastroenterol* 2009, **15**(47):5936-5945.
37. Furet JP, Kong LC, Tap J, Poitou C, Basdevant A, Bouillot JL, Mariat D, Corthier G, Dore J, Henegar C *et al*: **Differential adaptation of human gut microbiota to bariatric surgery-induced weight loss: links with metabolic and low-grade inflammation markers**. *Diabetes* 2010, **59**(12):3049-3057.
38. Schwiertz A, Taras D, Schafer K, Beijer S, Bos NA, Donus C, Hardt PD: **Microbiota and SCFA in lean and overweight healthy subjects**. *Obesity (Silver Spring)* 2010, **18**(1):190-195.
39. Balamurugan R, George G, Kabeerdoss J, Hepsiba J, Chandragunasekaran AM, Ramakrishna BS: **Quantitative differences in intestinal Faecalibacterium prausnitzii in obese Indian children**. *The British journal of nutrition* 2010, **103**(3):335-338.
40. Million M, Maraninchi M, Henry M, Armougom F, Richet H, Carrieri P, Valero R, Raccah D, Vialettes B, Raoult D: **Obesity-associated gut microbiota is enriched in Lactobacillus reuteri and depleted in Bifidobacterium animalis and Methanobrevibacter smithii**. *Int J Obes (Lond)* 2012, **36**(6):817-825.
41. Armougom F, Henry M, Vialettes B, Raccah D, Raoult D: **Monitoring bacterial community of human gut microbiota reveals an increase in Lactobacillus in obese patients and Methanogens in anorexic patients**. *PloS one* 2009, **4**(9):e7125.
42. Noor SO, Ridgway K, Scovell L, Kemsley EK, Lund EK, Jamieson C, Johnson IT, Narbad A: **Ulcerative colitis and irritable bowel patients exhibit distinct abnormalities of the gut microbiota**. *BMC Gastroenterol* 2010, **10**:134.
43. Carroll IM, Ringel-Kulka T, Siddle JP, Ringel Y: **Alterations in composition and diversity of the intestinal microbiota in patients with diarrhea-predominant irritable bowel syndrome**. *Neurogastroenterology and motility : the official journal of the European Gastrointestinal Motility Society* 2012, **24**(6):521-530, e248.
44. Duboc H, Rainteau D, Rajca S, Humbert L, Farabos D, Maubert M, Grondin V, Jouet P, Bouhassira D, Seksik P *et al*: **Increase in fecal primary bile acids and dysbiosis in patients with diarrhea-predominant irritable bowel syndrome**. *Neurogastroenterology*

and motility : the official journal of the European Gastrointestinal Motility Society 2012, **24**(6):513-520, e246-517.
45. Kerckhoffs AP, Ben-Amor K, Samsom M, van der Rest ME, de Vogel J, Knol J, Akkermans LM: **Molecular analysis of faecal and duodenal samples reveals significantly higher prevalence and numbers of Pseudomonas aeruginosa in irritable bowel syndrome**. *Journal of medical microbiology* 2011, **60**(Pt 2):236-245.
46. Saulnier DM, Riehle K, Mistretta TA, Diaz MA, Mandal D, Raza S, Weidler EM, Qin X, Coarfa C, Milosavljevic A *et al*: **Gastrointestinal microbiome signatures of pediatric patients with irritable bowel syndrome**. *Gastroenterology* 2011, **141**(5):1782-1791.
47. Rajilic-Stojanovic M, Biagi E, Heilig HG, Kajander K, Kekkonen RA, Tims S, de Vos WM: **Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome**. *Gastroenterology* 2011, **141**(5):1792-1801.
48. Sokol H, Seksik P, Furet JP, Firmesse O, Nion-Larmurier I, Beaugerie L, Cosnes J, Corthier G, Marteau P, Dore J: **Low counts of Faecalibacterium prausnitzii in colitis microbiota**. *Inflamm Bowel Dis* 2009, **15**(8):1183-1189.
49. Andoh A, Imaeda H, Aomatsu T, Inatomi O, Bamba S, Sasaki M, Saito Y, Tsujikawa T, Fujiyama Y: **Comparison of the fecal microbiota profiles between ulcerative colitis and Crohn's disease using terminal restriction fragment length polymorphism analysis**. *J Gastroenterol* 2011, **46**(4):479-486.
50. Sokol H, Lepage P, Seksik P, Dore J, Marteau P: **Temperature gradient gel electrophoresis of fecal 16S rRNA reveals active Escherichia coli in the microbiota of patients with ulcerative colitis**. *Journal of Clinical Microbiology* 2006, **44**(9):3172-3177.
51. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, Jia W, Cai S, Zhao L: **Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers**. *ISME J* 2012, **6**(2):320-329.
52. Moore WE, Moore LH: **Intestinal floras of populations that have a high risk of colon cancer**. *Applied and environmental microbiology* 1995, **61**(9):3202-3207.
53. Shen XJ, Rawls JF, Randall T, Burcal L, Mpande CN, Jenkins N, Jovov B, Abdo Z, Sandler RS, Keku TO: **Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas**. *Gut Microbes* 2010, **1**(3):138-147.
54. Lozupone C, Hamady M, Knight R: **UniFrac--an online tool for comparing microbial community diversity in a phylogenetic context**. *BMC bioinformatics* 2006, **7**:371.
55. Bokulich NA, Joseph CM, Allen G, Benson AK, Mills DA: **Next-generation sequencing reveals significant bacterial diversity of botrytized wine**. *PloS one* 2012, **7**(5):e36357.
56. Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP, O'Toole PW: **Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions**. *Nucleic acids research* 2010, **38**(22):e200.
57. Russmann H, Kempf VA, Koletzko S, Heesemann J, Autenrieth IB: **Comparison of fluorescent in situ hybridization and conventional culturing for detection of Helicobacter pylori in gastric biopsy specimens**. *J Clin Microbiol* 2001, **39**(1):304-308.
58. Barber JJ, Grichting WL: **Australia's media campaign against drug abuse**. *Int J Addict* 1990, **25**(6):693-708.
59. Rigsbee L, Agans R, Foy BD, Paliy O: **Optimizing the analysis of human intestinal microbiota with phylogenetic microarray**. *FEMS Microbiol Ecol* 2011, **75**(2):332-342.