Systematic and Applied Microbiology 38 (2015) 442–452

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

Crohn associated microbial communities associated to colonic mucosal biopsies in patients of the western Mediterranean

Roberto Vidal[a,1], Daniel Ginard[b,1], Sam Khorrami[b], Merit Mora-Ruiz[c], Raul Munoz[c], Marcela Hermoso[a], Sara Díaz[c], Ana Cifuentes[c], Alejandro Orfila[c], Ramon Rosselló-Móra[c,*]

[a] Institute of Biomedical Sciences, Faculty of Medicine, Universidad de Chile, Santiago, CL 8380453, Chile
[b] Department of Gastroenterology and Palma Health Research Institute, Hospital Universitario Son Espases, Palma de Mallorca, Spain
[c] Institut Mediterrani d'Estudis Avançats (IMEDEA; CSIC-UIB), 07190 Illes Balears, Spain

ARTICLE INFO

Article history:
Received 27 February 2015
Received in revised form 13 June 2015
Accepted 15 June 2015

Keywords:
Crohn disease
Pyrotagging
Operational phylogenetic units
Gut microbiome
Firmicutes
Faecalibacterium prausnitzii

ABSTRACT

Next generation sequencing approaches allow the retrieval of several orders of magnitude larger numbers of amplified single sequences in 16S rRNA diversity surveys than classical methods. However, the sequences are only partial and thus lack sufficient resolution for a reliable identification. The OPU approach used here, based on a tandem combination of high quality 454 sequences (mean >500 nuc) applying strict OTU thresholds, and phylogenetic inference based on parsimony additions to preexisting trees, seemed to improve the identification yields at the species and genus levels. A total of thirteen biopsies of Crohn-diagnosed patients (CD) and seven healthy controls (HC) were studied. In most of the cases (73%), sequences were affiliated to known species or genera and distinct microbial patterns could be distinguished among the CD subjects, with a common depletion of *Clostridia* and either an increased presence of *Bacteroidetes* (CD1) or an anomalous overrepresentation of *Proteobacteria* (CD2). *Faecalibacterium prausnitzii* presence was undetectable in CD, whereas *Bacteroides vulgatus–B. dorei* characterized HC and some CD groups. Altogether, the results showed that a microbial composition with predominance of *Clostridia* followed by *Bacteroidetes*, with *F. prausnitzii* and *B. vulgatus–B. dorei* as major key bacteria, characterized what could be considered a balanced structure in HC. The depletion of *Clostridia* seemed to be a common trait in CD.

Introduction

Microbial diversity measurements based on environmental 16S rRNA genes have permitted the recognition of the vast diversity of as yet uncultured microorganisms in environmental samples [56]. Next generation sequencing (NGS) techniques produce orders of magnitude higher numbers of sequences than the conventional techniques used in microbial molecular ecology, and they have been thoroughly used to reveal diversity in environmental samples (e.g. [4,9,11,31,47,54]). In general, the identification of environmentally occurring discrete 16S rRNA groups (often assumed to be species) had been based on clustering sequences into operational taxonomic units (OTUs [35]) using an identity threshold of 97% [4,11,31,47] or even lower (95%; [54]). The 97% threshold has been commonly used in microbial molecular ecology to circumscribe putative species based on the observation made by Stackebrandt and Goebel [42]. However, for taxonomic purposes this value was later raised to 98.7% [41], which was considered to be a more adequate minimum threshold for this category [56], although it had not been readily implemented in the molecular ecology of prokaryotes. Embracing sequences at 97% would be too conservative and can lead to underestimation of diversity, since distinct species of the same genus may cluster together. However, the major pitfall of the NGS approach is that the length of the sequences is too short for taxonomic identifications at the species level [56], and the almost complete sequence of the 16S rRNA gene would still be desirable. Generally, 454 pyrosequencing in the past has rendered sequence lengths of <300 nuc (e.g. [9,11,31,47,54]). However, the results obtained with such short sequences seemed to be robust enough to mirror the observations made with classical techniques

* Corresponding author at: Grup de Microbiologia Marina, Department of Ecology and Marine Resources, Institut Mediterrani d'Estudis Avançats (CSIC-UIB), C/Miquel Marqués 21, 07190 Esporles, Illes Balears, Spain. Tel.: +34 971 611 826; fax: +34 971 611 761.
E-mail address: rossello-mora@uib.es (R. Rosselló-Móra).
[1] Both authors contributed equally.

http://dx.doi.org/10.1016/j.syapm.2015.06.008
0723-2020/

[11,31]. Lately, improvements in the methodology have allowed much longer amplicon stretches (up to 800 nucleotides with means of >550 nucleotides; [26]), which has led to a more robust identification power due to the larger information content. However, due to the higher costs of 454 pyrosequencing, Illumina is becoming the method of choice for most of the new studies on environmental microbial diversity, despite some evidence that both the method and the different regions studied may not produce robust results comparable to conventional methods [4]. For these reasons, among the different NGS techniques, 454 may still be the best choice for amplicon analyses due to the larger stretches produced with relative low error rates [20].

The human gut microbiota harbors one order of magnitude more cells than that of the somatic and germ cells of the complete body [1], with its complex communities actively participating in host homeostasis [12] and playing important roles in nutrient metabolism and protection [39]. The gut epithelium is covered by a protective mucus layer, which can be damaged in inflammatory bowel diseases (IBD) inducing a disintegration of the web structure [45], finally leading to a modification of the microbial communities with implications for host health. Among the IBD, Crohn disease (CD) is a chronic disorder characterized by patchy inflammation of the gastrointestinal tract. The specific causes remain unclear, but clinical and experimental data consider it to be a consequence of uncontrolled intestinal inflammation in response to elusive environmental enteric microbiota, and immuno-regulatory factors in genetically susceptible individuals [3]. The major consequences are inflammation and ulceration of the intestinal lining [25], with an important reduction of the mucus layer and important changes in the colonic microbiomes [45].

In this work, we present the study of the microbiota attached to the intestinal mucosa of 13 CD biopsies together with 7 healthy control biopsies (HC) by means of the 454 sequencing approach of high quality amplified 16S rRNA gene sequences (mean >550 nuc). In addition, an identification approach was applied based on the recognition of the operational phylogenetic units (OPUs) that are based on phylogenetic inference instead of sequence identity clustering, in order to reflect the diversity observed better [9,26]. The work aimed to evaluate the use of the OPU approach based on 454 technology in order to increase the resolution power at lower taxonomic categories (i.e. genus and species) and compare the associated microbiomes with the colon tissues of HC and CD.

Materials and methods

Patients and samples

Patients with a Crohn disease diagnosis that had to be submitted to colonoscopy for medical revision were recruited for the study between August 2011 and March 2012 (Table 1; control individuals were identified with an S and CD with a C following their collection number). Control subjects were individuals submitted to colonoscopy due to constipation, colorectal cancer screening or anal bleeding, with normal colonoscopy inspection and where inflammatory or other bowel disorders were discarded. None of the subjects were exposed to antibiotics for 1 month prior to colonoscopy, and they all signed an informed consent. Demographics and clinical characteristics (disease localization, inflammatory activity, behavior, medication and surgical history) were collected (Table 1). For practical reasons, most of the CD patients selected had previously been ileal or ileocecal resected as they were colonoscopied for routine analyses, and a few non-resected patients were also included. In all cases, colectomies or ileostomies were exclusion criteria. The CD biopsies were taken from colon intestines with a healthy aspect and avoiding ulceration or lesions in the mucosa. Subjects received standard bowel cleansing with a polyethylene glycol preparation the evening before colonoscopy. Biopsies of approximately 1 mg and 2 mm^3 were collected from the colon using biopsy forceps during the colonoscopy. Biopsies were immediately placed in sterile tubes and stored at −80 °C for DNA/RNA extraction. The Balearic Islands' Ethical Committee approved the study.

DNA extraction, PCR amplification and pyrosequencing

Total DNA was extracted using the FastDNA® SPIN Kit (MP Biomedicals), following the manufacturer's indications, and a FastPrep®-24 (MP Biomedicals) instrument for homogenization. DNA was quantified using a NanoDrop® Spectrophotometer (Thermo Fisher Scientific Inc.), then stored at −20 °C and 20 $\mu g\,\mu L^{-1}$. For amplification, bacterial primers (forward GM3 5′-AGAGTTTGATCMTGGC-3′ and reverse 907RM 5′-CCGTCAATTCMTTTGAGTTT-3′ [27,28] were used in a 30-cycle PCR at 50 °C (Tm). For pyrosequencing, a secondary PCR incorporated tags and linkers into the amplicons using a 1:10 dilution of the original products as template in a 5-cycle PCR with the same conditions. Primers GM3-PS and a variant of 907-PS were used for *Bacteria* (Supplementary Table S1). Products were purified using MSB® Spin PCRapace (INVITEK), following the manufacturer's instructions. The concentration of the barcoded-amplicons was measured with a NanoDrop and, finally, an equimolar mixture of the amplicons was sent to a sequencing company (LifeTechnologies, Valencia, Spain; or Macrogen Inc., Seoul, Korea). The samples were sequenced using 454 GS-FLX+ Titanium technology. The set of sequences has been deposited at the ENA sequence repository under the project accession numbers PRJEB6107 and ERP005574.

Sequence trimming, chimera check, OTU (operational taxonomic unit) clustering

Data was processed using the Mothur pipeline [38]. Briefly, low-quality sequences were removed (<300 nuc with a window size and average quality score of 25, no ambiguities and no mismatches in reads with primers and barcodes were allowed, and a maximum homopolymer of 8 nucleotides). The 10-nuc barcode was used for sample identification. Chimeras were removed with the application Chimera Uchime implemented in Mothur. Sequences were clustered into OTUs (i.e. unique sets of sequences clustered equal to or above a certain identity threshold) at 99% using the UCLUST tool included in QIIME [6]. The most abundant read for each OTU was selected as representative.

Phylogenetic affiliation and OPU (operational phylogenetic unit) design

OTU representatives were added to the non-redundant SILVA REF111 database using the ARB program package [22,33]. Sequences were aligned with SINA using the LTP111 database as a template [32,55]. Alignments were manually inspected and improved, and sequences were added with the ARB parsimony tool to a default tree containing approximately 244,000 sequences of *Bacteria*. The closest relative sequences of an acceptable quality (almost full length with low indetermination and homopolymer occurrence) were selected and merged with the LTP111. The neighbor-joining algorithm was used for the final tree reconstruction. OTU representatives were added to the final tree with the parsimony tool. Sequences were grouped in OPUs based on the manual inspection of the tree [50]. An OPU was considered as the smallest clade containing one or more amplified sequences affiliating together with reference sequences available in the public repositories. When possible, the OPUs should include a type strain sequence present in the LTP database [55], and for identity values