



# Nasopharyngeal Microbiota in Children With Invasive Pneumococcal Disease: Identification of Bacteria With Potential Disease-Promoting and Protective Effects

## OPEN ACCESS

### Edited by:

Learn-Han Lee,  
Monash University Malaysia, Malaysia

### Reviewed by:

Emmanuel F. Mongodin,  
University of Maryland, Baltimore,  
United States  
Nuria Salazar,  
Instituto de Productos Lácteos de  
Asturias (IPLA), Spain

### \*Correspondence:

Alex Mira  
mira\_ale@gva.es  
Carmen Muñoz-Almagro  
cma@sjdhospitalbarcelona.org

†These authors have contributed  
equally to this work and are joint  
senior authors

‡Membership of the Catalan Study  
group of host-pathogen interaction in  
patients with IPD is provided in the  
Acknowledgments

### Specialty section:

This article was submitted to  
Microbial Symbioses,  
a section of the journal  
Frontiers in Microbiology

Received: 03 April 2018

Accepted: 07 January 2019

Published: 28 January 2019

### Citation:

Camelo-Castillo A, Henares D,  
Brotos P, Galiana A, Rodriguez JC,  
Mira A and Muñoz-Almagro C (2019)  
Nasopharyngeal Microbiota in  
Children With Invasive Pneumococcal  
Disease: Identification of Bacteria With  
Potential Disease-Promoting and  
Protective Effects.  
Front. Microbiol. 10:11.  
doi: 10.3389/fmicb.2019.00011

Anny Camelo-Castillo<sup>1</sup>, Desirée Henares<sup>1,2</sup>, Pedro Brotos<sup>1,2</sup>, Antonio Galiana<sup>3</sup>,  
Juan Carlos Rodríguez<sup>3</sup>, Alex Mira<sup>2,4\*†</sup> and  
Carmen Muñoz-Almagro<sup>1,2,5\*†</sup> on behalf of the Catalan Study Group of Host-  
Pathogen Interaction in Patients With IPD<sup>‡</sup>

<sup>1</sup> Institut de Recerca Sant Joan de Déu, Hospital Sant Joan de Déu, Barcelona, Spain, <sup>2</sup> Network of Epidemiology and Public Health, CIBERESP, Barcelona, Spain, <sup>3</sup> Department of Microbiology, Hospital de Alicante, Alicante, Spain, <sup>4</sup> Center for Advanced Research in Public Health, FISABIO Foundation, Valencia, Spain, <sup>5</sup> School of Medicine, Universitat Internacional de Catalunya, Barcelona, Spain

**Background and Aims:** The risk of suffering from some infectious diseases can be related to specific microbiota profiles. Specifically, the nasopharyngeal microbiota could play a role as a risk or protective factor in the development of invasive disease caused by *S. pneumoniae*.

**Methodology:** We analyzed the nasopharyngeal microbiota of children with invasive pneumococcal disease (IPD) and that of healthy controls matched by age, sex, and seasonality from Catalonia, Spain. Epidemiological, microbiological and clinical variables were considered to compare microbiota profiles, analyzed by sequencing the V1–V4 region of the 16S rRNA gene.

**Results:** Twenty-eight children with IPD (median age 43 months) and 28 controls (42.6 months) were included in the study. IPD children presented a significantly higher bacterial diversity and richness ( $p < 0.001$ ). Principal coordinate analysis revealed three different microbiota profiles: microbiota A, dominated by the genus *Dolosigranulum* (44.3%); Microbiota B, mostly represented by *Streptococcus* (36.9%) and *Staphylococcus* (21.3%) and a high diversity of anaerobic genera including *Veillonella*, *Prevotella* and *Porphyromonas*; and Microbiota C, mainly containing *Haemophilus* (52.1%) and *Moraxella* (31.4%). The only explanatory factor for the three microbiotas was the classification of children into disease or healthy controls ( $p = 0.006$ ). A significant negative correlation was found between *Dolosigranulum* vs. *Streptococcus* ( $p = 0.029$ ), suggesting a potential antagonistic effect against pneumococcal pathogens.

**Conclusions:** The higher bacterial diversity and richness in children with IPD could suggest an impaired immune response. This lack of immune competence could be aggravated by breastfeeding <6 months and by the presence of keystone pathogens

such as *Porphyromonas*, a bacterium which has been shown to be able to manipulate the immune response, and that could favor the overgrowth of many proteolytic anaerobic organisms giving rise to a dramatic dysbiosis. From an applied viewpoint, we found suggestive microbiota profiles associated to IPD or asymptomatic colonization that could be used as disease biomarkers or to pave the way for characterizing health-associated inhabitants of the respiratory tract. The identification of beneficial bacteria could be useful to prevent pneumococcal infections by integrating those microorganisms in a probiotic formula. The present study suggests not only respiratory tract samples, but also breast milk, as a potential source of those beneficial bacteria.

**Keywords:** nasopharyngeal microbiota, children, invasive pneumococcal disease, *Streptococcus pneumoniae*, *Dolosigranulum*

## INTRODUCTION

The human microbiota communities and their genes are an intriguing ecosystem that play an essential role for human body functions including food digestion, nutrition, regulation of human metabolism, and regulation of immune defense against infections, among others (Zhu et al., 2010). The majority of microorganisms of the human ecosystem act as symbionts that co-evolve and co-adapt with their human host according to their diet, life-style, human genetic characteristics, immune modulation, or environment parameters (Dethlefsen et al., 2007). Particularly in samples from the respiratory tract, the variability in bacterial composition between individuals has been described to be so high that a core microbiome could not be defined at the species level (Bogaert et al., 2011). Despite this inter-individual variability, the bacterial composition of each person has been proposed to fall within discrete categories where some bacterial taxa dominate the community. The presence of these structured microbial consortia was first described in the gut, where three groups or “enterotypes” were found, dominated by *Bacteroides*, *Prevotella*, or *Ruminococcus* (Arumugam et al., 2011). However, the notion of enterotypes is currently under debate, and it is still not clear whether enterotypes can be generalized to the general population, or whether microbiota follows in fact a gradient in taxonomic composition (Jeffery et al., 2012). In children between 1 and 3 months of age, for instance, the hypopharynx was found to contain four or five community types, and these “pneumotypes” followed specific trajectories during the child’s development (Mortensen et al., 2016). In adults, bronchoalveolar lavage samples showed two distinct pneumotypes, one of which was enriched with oral microorganisms (particularly *Prevotella* and *Veillonella*) and appeared to be associated with higher inflammatory markers (Segal et al., 2013). Thus, studying the bacterial community structures in the respiratory tract could be important to understand susceptibility to Chronic Obstructive Pulmonary Disease (COPD) or to pulmonary infections like pneumococcal pneumonia or, more extensively, Invasive Pneumococcal Disease (IPD).

It has been proposed that the degree of stability of the microbiota depends partly on the first colonization by keystone

species in the first days of life (Relman, 2012). Natural feed of humans in their early days is breastfeeding, so human milk is expected to be a major supplier of keystone species for human microbiota development, including the respiratory tract, given the high microbial diversity in colostrum, transition, and mature milk (Hunt et al., 2011; Cabrera-Rubio et al., 2012). In addition, breast milk provides macronutrients, micronutrients and bioactive molecules that protect against infections and inflammation and are key factors for human growth and development (Ballard and Morrow, 2013). Another source of resilience to infection is provided by the antagonistic interactions between the local microbiota and the invasive species. For instance, the nasal passage inhabitant *Corynebacterium accolens* appears to be over-represented in children without *Streptococcus pneumoniae* nasal colonization and has been shown to release antipneumococcal free fatty acids from human skin surface triacylglycerols (Bomar et al., 2016). Other commensal bacteria produce antimicrobial peptides that have been shown to inhibit oral pathogens (López-López et al., 2017) and prevent pharynx infections in children (Di Pierro et al., 2014). Thus, understanding microbiota composition and structure in respiratory samples will be instrumental for establishing disease risk and for the development of probiotics that could be used to colonize the respiratory tract and contribute to the prevention of infections (Tagg and Dierksen, 2003; Medina et al., 2008).

Our hypothesis is that the nasopharyngeal microbiota plays a role as a risk or protective factor in the development of invasive disease caused by *S. pneumoniae*. To test this hypothesis, the objective of the present study was to characterize the nasopharyngeal microbiota profiles in two groups of children: (i) children with invasive pneumococcal disease (IPD), considered as a case group whose nasopharyngeal microbiota was suffering an important disturbance; and (ii) a matched control group of healthy children representative of a healthy nasopharyngeal niche. A series of epidemiological, microbiological, and clinical variables related with major risk of developing IPD were considered to compare microbiota profiles in the two groups. For robust characterization of bacterial composition, a long-fragment 16S sequencing approach was used, in which a 780 bp region of the gene was

amplified and sequenced in order to improve taxonomic assignment.

## MATERIALS AND METHODS

### Study Design

This was an observational prospective study including all children and adolescents (<18 years) with IPD admitted to 5 hospitals in Catalonia, Spain: Hospital Sant Joan de Deu, Hospital de Nens, Hospital de Mataro, Hospital de Vic and Hospital de Calella (HC), during the period January 2014 to March 2015. One healthy control that was treated at Sant Joan de Deu hospital for minor surgery (i.e., phimosis or minor dermatologic surgery) was selected for each case. Cases and controls were matched by age, sex, and seasonality, given previous evidence of seasonal influence on the microbiota composition of respiratory samples (Bogaert et al., 2011). Children whose parents or guardians did not sign the informed consent were excluded of the study. The study was performed following the guidelines of the Ethics Committee of Hospital Sant Joan de Deu which approved the study. All information collected has been treated confidentially and in accordance with applicable laws on personal data.

### IPD Definition

IPD was defined as the presence of clinical findings of infection (which were used for classification of the disease) together with isolation of *Streptococcus pneumoniae* and/or DNA detection of autolysin (*LytA*) gene and an additional capsular gene of *S. pneumoniae* by Real-Time PCR in plasma, cerebrospinal fluid or pleural fluid. Clinical diagnoses were mutually exclusive.

### Data Collection

Relevant clinical, epidemiological, immunological, and microbiological variables were collected for each subject including delivery mode, duration of breastfeeding (mixed or exclusive), fulfillment of breastfeeding world health recommendations (minimum duration of 6 months) exposure to crowding, pneumococcal vaccination status, parents' smoking habits, and previous occurrence of respiratory infections. Among microbiological variables we assessed carriers of 13-valent pneumococcal conjugate vaccine (PCV13) serotypes, invasiveness potential of nasopharyngeal *S. pneumoniae* serotypes and co-colonization with respiratory viruses. Serotypes were classified according to the studies of Brueggemann et al. (2003), Sleeman et al. (2006), and del Amo et al. (2014). Serotypes 1, 3, 4, 5, 7F, 8, 9A, 9V, 12F, 14, 18C, and 19A were considered to have a high-attack rate or highly-invasive serotypes whereas the remainder were considered as low-attack rate or non-highly-invasive or opportunistic serotypes. In addition, we recorded data of clinical diagnosis, presence of underlying disease, previous antibiotic intake, white blood cell count, C-reactive protein level, occurrence of complications, length of hospitalization, provision of intensive care, and course of disease for the cases.

## Sample Collection and Laboratory Analyses

Nasopharyngeal aspirate samples collected in sterile plastic tubes from cases and controls and eluted with 0.5 ml of phosphate buffered saline (PBS) were extracted and stored at  $-80^{\circ}\text{C}$  until further genomic/microbial analysis. Whenever samples were needed for the sole purposes of the study they were taken by experienced nurses of the Hospital Sant Joan de Deu Clinical Trials Unit.

## Microbiological Methods

### *S. pneumoniae* Detection and Characterization

The definition of Invasive Pneumococcal Disease (IPD) is the presence of clinical findings of infection together with isolation of *Streptococcus pneumoniae* by culture and/or DNA detection of *S. pneumoniae* by Real-Time PCR in plasma, cerebrospinal fluid or any other sterile fluid. The microbiological confirmation of the patients is based on one technique, the other or both together. All pneumococcal isolates were identified by standard microbiological methods, including the optochin sensitivity test and an antigenic test targeting the capsular polysaccharide (Slidex pneumo-kit, bioMérieux, Marcy-l'Étoile, France). DNA detection of pneumococcal *LytA* gene was performed by a duplex real-time PCR as previously reported (Selva et al., 2014; del Amo et al., 2015). This duplex real-time PCR also included as target the *Rnase P* human single-copy gene to detect and measure the number of human cells (Brotons et al., 2017), using primers and probes recommended by CDC (Meningitis, 2011). Absolute quantification of RNaseP using PCR can be used to estimate the number of human cells present in the clinical samples. The measure of Human DNA load was then used to normalize by potential differences in sample amount and avoid potential bias between samples. In addition, DNA was measured prior to PCR and concentrations adjusted to minimize amplification bias between samples. Capsular typing was carried out by a molecular technique based on automated fluorescent fragment analysis which allows differentiation of 40 serotypes (Selva et al., 2014). Quellung reaction performed at the National Center for Microbiology (Majadahonda, Madrid) was used to complete serotyping in invasive strains isolated by culture.

### Viral Respiratory Study in Nasopharyngeal Samples

All nasopharyngeal samples were tested by Anyplex™ II RV16 Detection v1.1. (Seegene, Seoul, Korea). This multiplex Real-Time PCR assay performs simultaneous amplification, detection and differentiation of DNA/RNA of Adenovirus (AdV), Influenza A virus (FluA), Influenza B virus (FluB), Parainfluenza virus 1 (PIV1), Parainfluenza virus 2 (PIV2), Parainfluenza virus 3 (PIV3), Parainfluenza virus 4 (PIV4) Rhinovirus A/B/C (HRV), Respiratory syncytial virus A (RSV A), Respiratory syncytial virus B (RSV B), Bocavirus 1/2/3/4 (HBoV), Metapneumovirus (MPV), Coronavirus 229E (CoV 229E), Coronavirus NL63 (CoV NL63), Coronavirus OC43 (CoV OC43), Enterovirus (HEV), and an internal control in two PCR sets.

## Microbiota Analysis

DNA extraction of nasopharyngeal aspirates was performed by Nuclisens® EasyMag® according to manufacturer instructions (bioMérieux, Marcy-l'Étoile, France). This is an automated system for total nucleic acid extraction based in magnetic silica particles. After extracting the DNA, its quality and quantity was measured with Nanodrop (Thermo Fisher Scientific, Massachusetts, USA) and only samples with a 260/280 absorbance ratio between 1.8 and 2 were processed for Next Generation Sequencing (NGS).

NGS libraries were created with 100 ng of total DNA for each sample, to which a unique multiplex identifier (MID) was assigned. PCR amplification of the 16S rRNA gene was performed using high fidelity Extensor Long Range PCR Enzyme (Thermo Fisher Scientific, Massachusetts, USA), with the degenerate universal bacterial primers of 16S rRNA gene 8F (5'-CAGAGTTTGATCMTGGCTCAG-3') and 785R (5'-GGCCVGGTATCTAATCC-3') (Simón-Soro et al., 2014) and the following cycling conditions to minimize amplification biases (Simón-Soro et al., 2014): 2 min at 94°C, followed by 30 cycles of 10s at 94°C, 30s at 52°C, 90s at 68°C, and a final step of 7 min at 68°C. These primers amplified the variable regions 1 to 4 (V1–V4) of 16S rRNA gene with an expected size of 780 pb. The PCR products were first purified using a Minielute PCR purification Kit (Qiagen, Venlo, The Netherlands) and then using Agencourt AMPure beads (Beckman coulter, Munich, Germany). Finally, the amplicons were pyrosequenced using a 454 GS FLX Titanium chemistry (Roche, Basel, Switzerland) with Lib-L type microspheres, pooling 20 samples per 1/8 of a pyrosequencing plate.

Negative controls (PBS) were extracted by the same method of the studied samples (Easymag) and 16S rRNA gene and Human *Rnase P* gene were analyzed by PCR, producing no amplification. DNA concentrations for negative controls were also below the quality control thresholds required at Macrogen Inc. (Republic of South Korea). In addition, a negative control (water) was included in each sequencing run to discard contamination during library preparation, producing no results.

## Bioinformatic Analysis and OTU Assignment

Raw sequences were analyzed using “Quantitative Insights into Microbial Ecology” (QIIME) v1.7.0 software (Caporaso et al., 2010) and were separated using the 8-bp “barcodes” assigned to each sample. Chimeric sequences were eliminated by ChimeraSlayer program (Haas et al., 2011). An end-trimming quality filtering was performed by removing 50-bp windows with quality values <25 using Prinseq. Only reads >500 bp were included for taxonomic classification. Sequences with differences in the primer binding region or with more than four ambiguities in homopolymeric regions were excluded from the analysis. The resulting high quality sequences of the 16S rRNA gene were classified using the RDP database (Ribosomal Database Project) (Cole et al., 2009) with a bootstrap cutoff of 80%. Samples were classified in Operational Taxonomic Units (OTUs) at 97% sequence identity as the standard species-level boundary

(Yarza et al., 2008). Only the OTUs representing over 0.1% of the total sequences of each sample were considered for further statistical analysis, as low-frequency reads, including singletons, are more likely to represent sequencing errors, contaminants, or transient organisms without a biological role at the niche under study (de la Cuesta-Zuluaga and Escobar, 2016). Bacterial taxonomic composition was determined for each sample and means for each group (cases and controls) were calculated.

## Estimation of Microbial Diversity and Statistical Analysis

Data analysis was performed using QIIME 1.7.0 software. Rarefaction curves (Mean ± S.E.) were calculated by including 1,000 randomly selected reads per sample. Alpha diversity indexes were calculated from rarefied samples (1,000 sequences per sample), using the Shannon index (Shannon, 1948) for diversity and the Chao1 index (Chao, 1984) for richness. Beta diversity was also calculated with the rarefied 1,000 reads per sample using weighted and unweighted UniFrac (Lozupone et al., 2006) distance matrices, and the principal coordinate analysis (PCoA) generated 2D and 3D plots for all mapping fields. In addition, we clustered samples using UPGMA (Unweighted Pair Group Method with Arithmetic mean, also known as average linkage) (Michener and Sokal, 1957). Two-way comparisons in bacterial composition using the UniFrac metric (Lozupone et al., 2006) were used to measure whether the microbial communities in the microbiota types were significantly different. A tree with 1,000 reads from each sample was obtained, and microbiota types were considered significantly different if the UniFrac distance value for the tree was larger than expected if the sequences were randomly distributed. A thousand permutations were performed to obtain a *P*-value, using Bonferroni corrections for multiple comparisons. In addition, we also performed Constrained Correspondence Analysis (CCA), which is a statistic tool which emphasizes variation, and tests whether the factor provided (the microbiota type) can explain part of the total variability. This analysis was performed by the R software ade4 package (Dray and Dufour, 2007) using the Chi-squared distances-based function CCA. Adonis tests were done with the R library “vegan” (Oksanen et al., 2018) to determine corrected *p*-values.

Microbial community comparisons and statistical analysis were performed using the statistical software R v2.15.2 with the packages for Community Ecology (vegan), Euclidean Methods in Environmental Sciences (ade4), The Bee Swarm Plot (beeswarm) and gplots packages (<http://R-project.org>), R (The R Foundation for Statistical Computing).

We used LEfSe (Linear discriminant analysis Effect Size) with default parameters (Segata et al., 2011), available in the Galaxy Web Server toolkit to determine significant differences (alpha value = 0.05) in the proportions of bacterial genera between cases and controls groups. LEfSe uses the non-parametric factorial Kruskal-Wallis (KW) sum-rank test to detect features with significant differential abundance with respect to the class of interest.

Normality of the data was analyzed with the Shapiro-Wilk test. Continuous data with normal distributions were described using mean and Standard Deviation (SD). In case of non-normal distribution, median and interquartile range (IQR) were used. Significance ( $p$ -values) of continuous and normally distributed data was determined using Student's  $t$ -test. When data did not follow a normal distribution, Wilcoxon signed-rank test was used. Significance of categorical data was established by using Chi-square test. When Chi-square expected frequency was equal or less than five, Fisher's exact test was applied. When variables followed a normal distribution, the ANOVA test was used to determine if there were statistically significant differences in mean values between the microbiota groups. If distribution

was not normal, we used the non-parametric Kruskal-Wallis test. Multiple testing correction was performed when needed.

## Ethics Statement

The study was performed following the guidelines of the Ethics Committee of Hospital Sant Joan de Deu which approved the study (project approval code PIC-70-15). All information collected has been treated confidentially and in accordance with applicable laws on personal data. Written consent was obtained from parent or legal guardians of children included in the study.

**TABLE 1** | Epidemiological and microbiological characteristics of study groups.

Variable	Cases ( $n = 28$ )	Controls ( $n = 28$ )	$p$ -value
<b>EPIDEMIOLOGICAL VARIABLES</b>			
Age, months (Md, IQR)	43.0 (20.8–60.2)	42.6 (31.6–58.9)	0.72
Gender, male	16 (57.1)	16 (57.1)	1.00
Influenza season	12 (42.9)	14 (50.0)	0.59
Race, Caucasian	23 (82.1)	25 (89.3)	0.71
Birth weight, gr (Mean, SD)	3,143 (626)	3,298 (579)	0.34
Gestational age, weeks (Mean, SD)	38.5 (2.3)	38.8 (2.0)	0.67
Delivery type, C-section <sup>†</sup> (cases: $n = 25$ ; controls: $n = 26$ )	7 (28.0)	10 (38.5)	0.43
Breastfeeding	23 (82.1)	19 (67.9)	0.22
Breastfeeding duration $\geq 6$ months <sup>¶</sup> (cases: $n = 23$ ; controls: $n = 19$ )	9 (39.1)	13 (68.4)	0.06
Breastfeeding duration, months (Md, IQR) <sup>†</sup> (cases: $n = 23$ ; controls: 19)	4 (1–7)	6 (4–10)	0.19
Day-care attendance	25 (89.3)	23 (82.1)	0.45
Shared bedroom <sup>†</sup> (cases: $n = 21$ ; controls: 27)	13 (61.9)	12 (44.4)	0.23
Parental smoking <sup>†</sup> (cases: $n = 26$ )	11 (42.3)	13 (46.4)	0.76
Vaccinated against influenza	1 (3.6)	0 (0.0)	1.00
Vaccinated against pneumococcal disease	17 (60.7)	19 (67.9)	0.58
PCV13 vaccinated	14 (82.4)	17 (89.5)	0.42
Respiratory infection in the previous 30 days <sup>†</sup> (Controls: $n = 27$ )	13 (46.4)	13 (48.1)	0.90
<b>MICROBIOLOGICAL VARIABLES</b>			
Sp detection in NP sample by PCR	18 (64.3)	16 (57.1)	0.58
High-attack rate Sp serotype in NP sample	10 (55.5)	2 (12.5)	<b>0.009</b>
PCV13 serotype coverage in NP sample	11 (61.1)	2 (12.5)	<b>0.004</b>
Viral detection in NP sample by PCR	17 (60.7)	14 (50.0)	0.42
Rhinovirus	10 (35.7)	10 (35.7)	1.00
Adenovirus	4 (14.3)	4 (14.3)	1.00
Parainfluenza virus	2 (7.1)	1 (3.6)	1.00
Respiratory syncytial virus	2 (7.1)	0 (0.0)	0.49
Coronavirus	1 (3.6)	2 (7.1)	1.00
Enterovirus	1 (3.6)	0 (0.0)	1.00
Bocavirus	1 (3.6)	0 (0.0)	1.00
Influenza A virus	2 (7.1)	0 (0.0)	0.49
Influenza B virus	0 (0.0)	1 (3.6)	1.00
Multiple respiratory viral infection	5 (29.4)	4 (28.6)	1.00

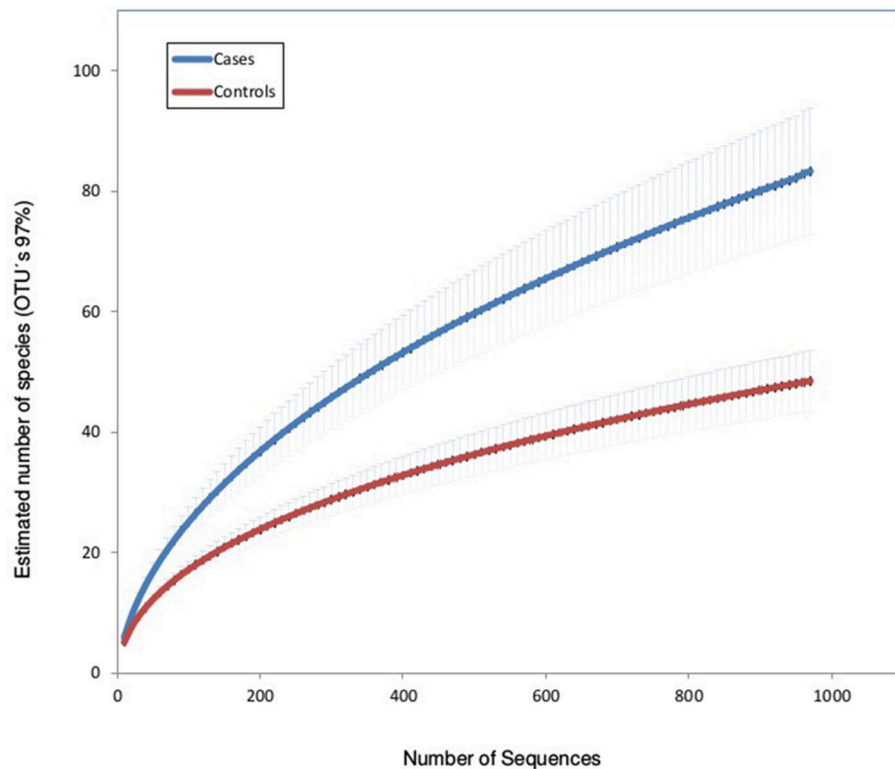
Results expressed as counts (%) unless stated otherwise.

Md, median; IQR, interquartile range; SD, standard deviation; PCV13, 13-valent pneumococcal conjugate vaccine; PICU, pediatric intensive care unit.

<sup>†</sup>missing values.

<sup>¶</sup>according to WHO recommendations.

The bold values means statistically significant associations with  $p$ -values  $< 0.05$ .



**FIGURE 1** | Rarefaction curves of nasopharyngeal samples in patients with invasive pneumococcal disease (cases) and healthy children (controls). The horizontal axis shows the number of reads (sequencing effort) obtained by pyrosequencing the 16S rRNA gene. The vertical axis shows the number (mean  $\pm$  S.E.) of operational taxonomic units (OTUs) at a level of 97% (estimated mean number of bacterial species).

## RESULTS

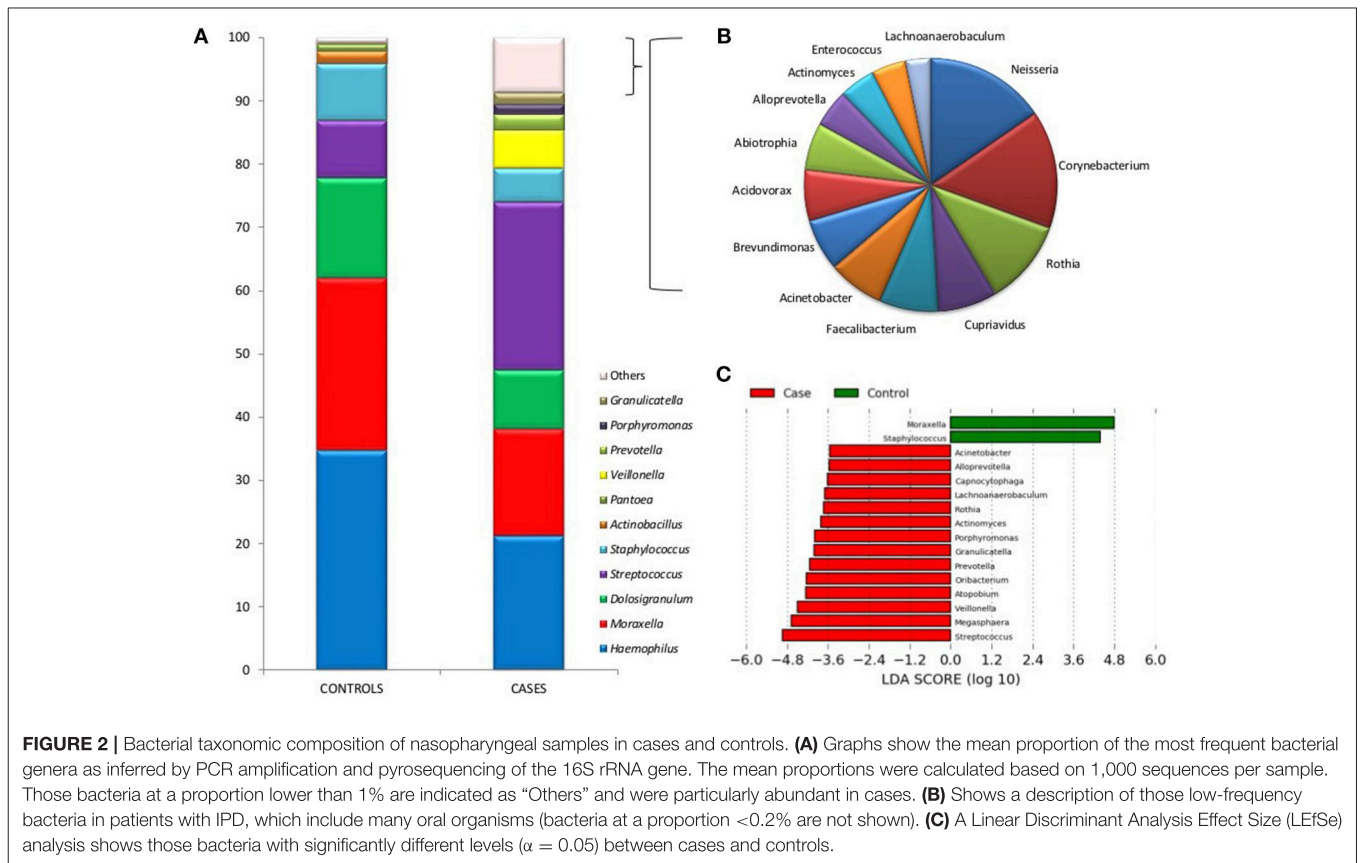
### Epidemiological and Clinical Features of Patients

During the study period a total of 53 patients were diagnosed of IPD in Hospital Sant Joan de Deu ( $n = 38$ ), Hospital de Nens ( $n = 6$ ), Hospital de Mataro ( $n = 5$ ), Hospital de Vic ( $n = 2$ ), and Hospital de Calella ( $n = 2$ ). Forty-two of them (79.2%) accepted to participate in the study under informed consent. Twenty-eight (52.8%) controls were available for matching with cases so the study finally comprised 56 subjects. No significant differences of total human DNA load was found between samples of cases and controls ( $\text{Log}_{10}$  3.6 vs. 3.2 gene copies/reaction,  $p = 0.1$ ), indicating that both types of samples did not differ in the amount of material.

The median age of the 28 cases was 43.0 months (IQR 20.8–60.2 months) and 16 (57.1%) were male. The clinical diagnoses of these patients were 19 pneumonia (including 10 complicated pneumonia with empyema and 3 necrotizing pneumonia), 7 bacteremia, 1 meningitis, and 1 fulminant sepsis. Invasive samples collected for diagnosing IPD included blood ( $n = 19$ , 67.8%), pleural fluid ( $n = 5$ , 17.9%), blood and pleural fluid ( $n = 3$ , 10.7%), and cerebrospinal fluid ( $n = 1$ , 3.6%). Detection of *S. pneumoniae* in invasive samples was performed by PCR ( $n = 14$ , 50.0%), by culture ( $n = 12$ , 42.9%), and both by PCR and culture ( $n = 2$ , 7.1%). Median white blood

cell count of 17,515/L and median C-reactive protein levels of 154.7 mg/L indicated inflammatory activity. An underlying disease was reported in 3 cases with chronic pulmonary disease, neuroblastoma, and splenic dysfunction, respectively. Median length of hospitalization was 7 days (IQR 4–8 days). Five cases (17.9%) received intensive care, 4 (14.3%) suffered sequelae, and one child died.

Pneumococcal nasopharyngeal carriage was observed in 18 (64.3%) cases and 16 (57.1%) controls. A total of 11 PCV13 serotypes were identified among carriers in the case group, specifically serotype 1 ( $n = 2$ ), serotype 3 ( $n = 3$ ), serotype 14 ( $n = 2$ ), serotype 19A ( $n = 3$ ), and serotype 19F ( $n = 1$ ). In contrast, only 2 PCV13 serotypes, classified as serotype 3, were found in control carriers. Nasopharyngeal samples were taken from cases after a mean of 92 h of fever (IQR 48–408 h) and all of them except two (one patient with necrotizing pneumonia and another one with fatal fulminant sepsis and viral respiratory coinfection), were exposed to beta-lactamic antibiotic treatment, during a mean period of 4 days. No patients were exposed to macrolide treatment or other class of antibiotics. **Table 1** shows epidemiological and microbiological variables for cases and controls. Proportions of coinfection with respiratory virus and pneumococcal nasopharyngeal carriage, were higher in cases compared to controls, but the differences between groups did not reach statistical significance. There was a significantly higher proportion of PCV13 pneumococcal serotype carriers



in cases compared to controls (61.1 vs. 12.5%,  $p = 0.004$ ) as well as significantly higher number of high-attack rate serotypes (55.5 vs. 12.5%,  $p = 0.009$ ). A trend for significance was observed for breastfeeding through 6 months (the minimum time recommended by World Health Organization) in cases (39.1 vs. 68.4%,  $p = 0.06$ ).

### High Bacterial Diversity in IPD Patients

A total of 284,558 16S rRNA sequence reads were obtained. After sequence length and quality filtering, we obtained 201,888 reads >500 bp (500–780 bp), with an average of 2,549 reads of the 16S rRNA gene per sample (range between 1,085 and 29,370 reads).

Rarefaction curves (using OTUs at 97% sequence identity) showed that mean bacterial richness had higher values in children with IPD compared with healthy patients (**Figure 1**). Individual rarefaction curves with 1,000 reads show that most bacteria present were detected by this level of coverage and show a shift toward higher OTU richness in children with IPD (**Supplementary Figure 1**). Species richness (Chao1 Index) was significantly higher in cases (mean: 79, 95% confidence interval: 57.4–100.58) compared to healthy controls (mean: 47.8, 95% CI: 37.5–58.2,  $p < 0.0001$ ). The Shannon diversity index was also significantly higher in cases (mean: 1.21, 95% CI: 0.94–1.48) vs. healthy controls (mean: 0.77, 95% CI: 0.62–0.92,  $p < 0.0001$ ).

The taxonomic assignment of the 16S rRNA reads revealed that the most common phyla in IPD patients were *Firmicutes* (50.9% of the total number of reads), *Proteobacteria* (41.4%),

*Bacteroidetes* (5.0%), *Actinobacteria* (2.2%), and a lower proportion of *Fusobacteria*. Within the control group, the most common phyla were *Proteobacteria* (66.0% of the total number of reads), *Firmicutes* (33.6%), *Bacteroidetes* (0.2%), and *Actinobacteria* (0.2%).

**Figure 2A** shows differences between the two groups in the mean proportions of each bacterial genera. In the cases group the most prevalent bacteria were *Streptococcus*, *Haemophilus*, *Moraxella*, *Dolosigranulum*, *Veillonella*, and *Staphylococcus* with 26.4, 21.1, 16.8, 9.3, 6.1, and 5.2% of the total, respectively. Whilst in the control group their proportions were 9.0, 34.5, 27.1, 15.7, 0.1, and 8.9%, respectively. The higher bacterial diversity detected in children with IPD corresponded to the high prevalence of lactate fermenting *Veillonella*, and mainly to genera found at low frequencies. The taxonomic assignment of those genera present at <1% of the total, which accounted for over 9% of the reads in IPD patients, is shown in **Figure 2B** and reveals the presence of many oral species like *Corynebacterium*, *Neisseria*, *Actinomyces*, or *Rothia*, among others.

LEfSE ranking analysis shows that *Moraxella* and *Staphylococcus* were found to be significantly more abundant in healthy individuals, according to this high-dimensional class comparison test. On the other hand, the genera *Streptococcus*, *Megasphaera*, *Veillonella*, *Atopobium*, *Oribacterium*, *Prevotella*, *Granulicatella*, *Porphyromonas*, *Actinomyces*, *Rothia*, *Lachnoanaerobaculum*, *Capnocytophaga*, *Alloprevotella*, and *Acinetobacter* were significantly more abundant in patients

**TABLE 2 |** Bacterial species identified in the present study.

Genus	Species	Cases (n = 28)	Controls (n = 28)
<i>Streptococcus</i>	<i>pneumoniae</i>	51.08%	91.17%
	<i>mitis</i>	23.10%	2.82%
	<i>oralis</i>	7.02%	0.51%
	sp	6.62%	1.07%
	uncultured sp	5.40%	1.49%
	<i>infantis</i>	3.19%	0.13%
	sp. oral clone	1.84%	0.64%
	<i>peroris</i>	0.74%	–
	<i>sanguinis</i>	0.41%	0.77%
	<i>parasanguinis</i>	0.22%	–
	<i>genomosp</i>	0.12%	–
	<i>pseudopneumoniae</i>	0.10%	–
	<i>cristatus</i>	0.07%	0.47%
	<i>pyogenes</i>	0.02%	0.94%
	<i>Dolosigranulum</i>	<i>pigrum</i>	83.26%
uncultured sp		16.74%	15.83%
<i>Haemophilus</i>	<i>influenzae</i>	80.30%	99.65%
	uncultured sp	11.77%	0.08%
	sp	6.81%	0.03%
	<i>parainfluenzae</i>	1.02%	0.09%
	<i>haemolyticus</i>	0.11%	0.10%
<i>Moraxella</i>	<i>catarrhalis</i>	94.89%	79.71%
	<i>lincolnii</i>	2.93%	8.30%
	<i>nonliquefaciens</i>	2.14%	4.46%
	<i>caprae</i>	0.04%	0.40%
	<i>lacunata</i>	–	7.14%
<i>Prevotella</i>	uncultured sp	34.54%	60.00%
	<i>histicola</i>	15.70%	–
	sp	12.80%	20.00%
	sp. oral clone	12.32%	–
	<i>pallens</i>	9.90%	–
	<i>nanceiensis</i>	3.86%	20.00%
	<i>conceptionens</i>	3.38%	–
	<i>oris</i>	2.90%	–
	<i>melaninogenic</i>	1.69%	–
	<i>salivae</i>	0.72%	–
<i>Veillonella</i>	uncultured sp	39.94%	24.00%
	<i>parvula</i>	29.49%	24.00%
	sp	14.23%	8.00%
	<i>dispar</i>	7.37%	24.00%
	sp. oral taxon	6.67%	20.00%
	<i>tobetsuensis</i>	2.24%	–
<i>Porphyromonas</i>	uncultured sp	71.93%	56.25%
	sp. oral clone	14.91%	18.75%
	sp	13.16%	25.00%
<i>Staphylococcus</i>	<i>epidermidis</i>	65.40%	9.67%
	sp	11.16%	47.12%
	<i>haemolyticus</i>	7.07%	0.05%
	uncultured sp	6.51%	2.23%

(Continued)

**TABLE 2 |** Continued

Genus	Species	Cases (n = 28)	Controls (n = 28)
	<i>pettenkoferi</i>	3,35%	–
	<i>devriesei</i>	2,51%	–
	<i>cohnii</i>	1,95%	–
	<i>hominis subsp. null</i>	1,40%	–
	<i>pasteuri</i>	0,28%	–
	<i>warneri</i>	0,28%	–
	<i>aureus</i>	–	40.89%

Data indicate percentages of 16S rRNA reads from a given genus with a hit against each species. Frequencies of each genera are shown in **Figure 2**. To minimize taxonomic assignment errors, only hits with >97% nucleotide identity and >500 bp alignment were selected. Only species with frequencies >0.1% are shown.

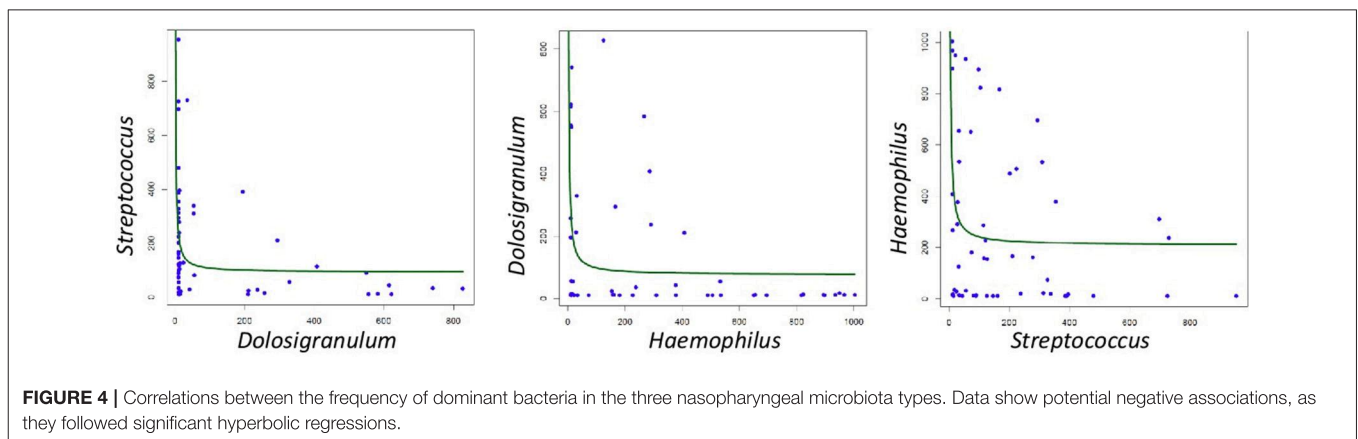
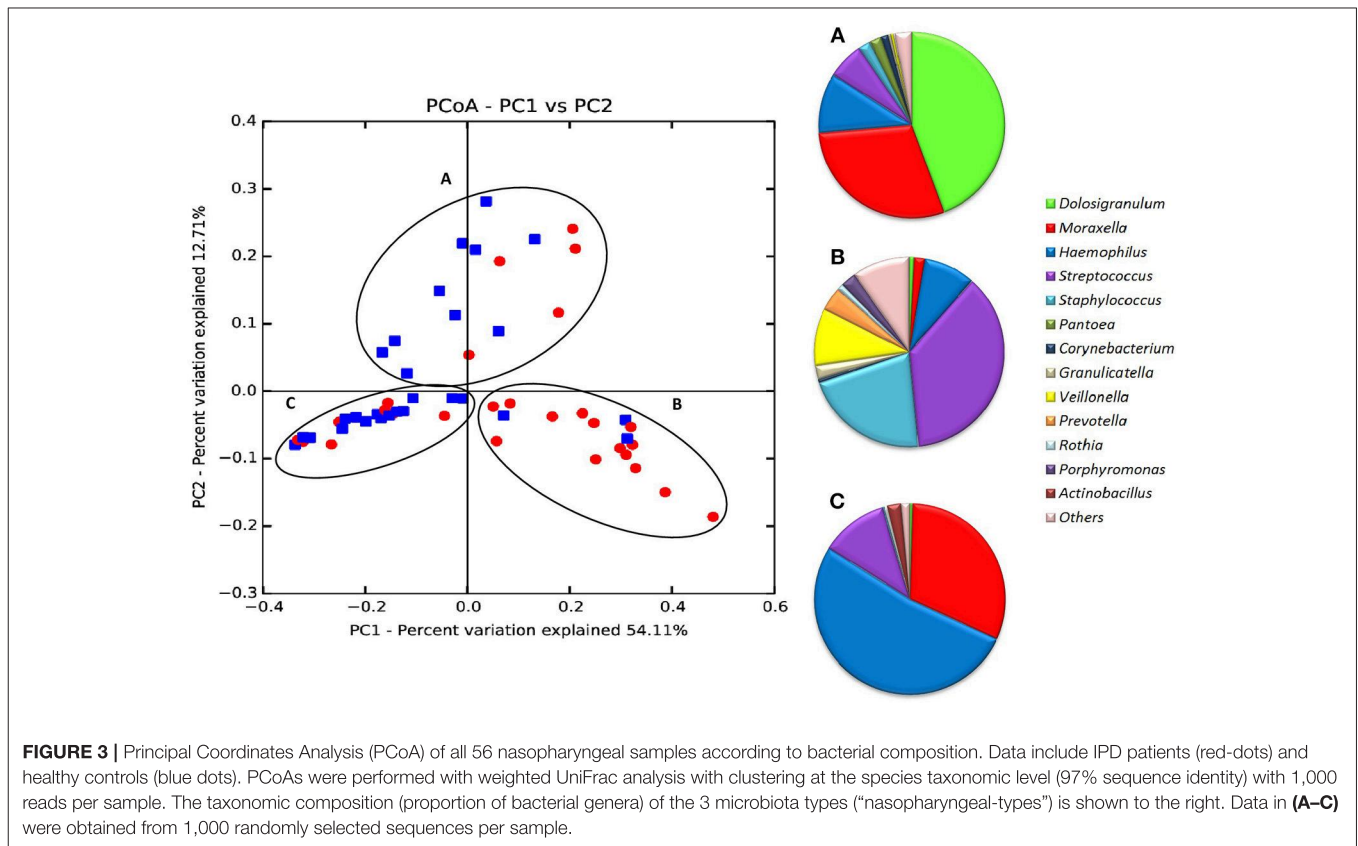
with IPD (**Figure 2C**). Most of these bacteria are common oral inhabitants (Mira et al., 2017). **Supplementary Figure 2** shows a Heatmap profile showing abundances of bacterial genera in individual cases and controls indicating the age of each child.

The use of long-amplicon (780 bp) sequencing allowed us to taxonomically assign the 16S sequences at the species level with a higher degree of accuracy (Shin et al., 2016). In order to minimize errors in taxonomic assignment, only sequence alignments >97% sequence identity over >500 bp were considered, which would encompass the 16S rRNA hypervariable regions v1–v3/v4. Our results are shown in **Table 2**. According to this assignment, the genus *Staphylococcus* would be dominated by *S. aureus* in the controls (>40% of the reads in this genus) whereas this species would be totally absent in IPD patients, that would be dominated by *S. epidermidis* and *S. haemolyticus*. Virtually, all *Haemophilus* species in controls would correspond to *H. influenzae*, whereas in the cases, 18% of the reads in this genus would correspond to uncultured or unknown species. It is surprising to note that over 90% of the reads matching *Streptococcus* would correspond to *S. pneumoniae* in the controls while in IPD patients, only 51% of the streptococcal reads would correspond to *S. pneumoniae* and over 40% of them gave top matches to *S. mitis*, *S. oralis*, or uncultured species.

## Microbiota Nasopharyngeal-Types in Health and Disease

When all samples were analyzed by two-dimensional principal coordinates analysis (2D PCoA), 67% of the variability in the data could be explained by the first two components. Samples appeared to cluster in three different groups, corresponding to three different types of respiratory tract microbiota, or nasopharyngeal-types (**Figure 3**). The Microbiota A was mainly composed of reads assigned to the genera *Dolosigranulum* (44.3%), *Moraxella* (29.3%), and *Haemophilus* (10.5%) (**Figure 3A**). The Microbiota B was represented mostly by the genera *Streptococcus* (36.9%), *Staphylococcus* (21.3%), *Veillonella* (9.8%), together with a high diversity of anaerobic genera as *Prevotella* and *Porphyromonas* (**Figure 3B**). Finally, the Microbiota C was composed mainly of the genera *Haemophilus* (52.1%), *Moraxella* (31.4%), and *Streptococcus*





(11.4%) (**Figure 3C**). Thus, nasopharyngeal-type A was dominated by *Dolosigranulum*, nasopharyngeal-type C by *Haemophilus*, and nasopharyngeal-type B by *Streptococcus*, with a high presence of oral microorganisms. The three microbiota types were found to be significantly different from each other (Unifrac Distance, corrected  $p < 0.01$  in all cases). In addition, CCA analysis showed that microbiota type significantly explained the variability in bacterial composition among samples (Adonis  $p$ -value: 0.002). In agreement with the existence of specific bacterial communities dominated by a given genus, negative correlations were found (**Figure 4**)

between the three dominant genera ( $p$ -values for significant hyperbolic regressions were, respectively, 0.029, 0.011, and 0.029 for comparisons between *Dolosigranulum* and *Streptococcus*, *Dolosigranulum* and *Haemophilus*, and *Streptococcus* and *Haemophilus*).

In order to understand the features influencing nasopharyngeal-types, all available epidemiological, microbiological, immunological and clinical variables were compared by bivariate analysis with the three microbiota types. No significant differences were found in any of the variables considered that could explain the grouping in the three different

**TABLE 3 |** Epidemiological, microbiological, immunological, and clinical characteristics by microbiota type.

<b>A</b>				
Variable	Microbiota A (n = 15)	Microbiota B (n = 17)	Microbiota C (n = 24)	p-value
Case group	5 (17.9)	14 (50.0)	9 (32.1)	<b>0.006</b>
<b>EPIDEMIOLOGICAL VARIABLES</b>				
Age, months (Md, IQR)	40.8 (20.8–80.3)	41.4 (32.1–56.6)	44.4 (29.5–61.4)	0.95
Gender, male	8 (53.3)	8 (47.1)	16 (66.7)	0.43
Influenza season	6 (40.0)	6 (35.3)	14 (58.3)	0.29
Race, Caucasian	11 (73.3)	16 (94.1)	21 (87.5)	0.26
Birth weight, gr (Mean, SD)	3,147 (689)	3,329 (749)	3,198 (428)	0.68
Gestational age, weeks (Mean, SD)	38.1 (2.3)	38.6 (2.6)	39.7 (1.7)	0.43
Delivery type, C-section <sup>†</sup> (n = 48)	6 (42.9)	4 (28.6)	7 (30.4)	0.75
Breastfeeding	13 (86.7)	13 (76.5)	16 (66.7)	0.43
Breastfeeding in controls	9 (90.0)	1 (33.3)	9 (60.0)	0.10
Breastfeeding duration ≥ 6 months <sup>¶</sup> † (A: n = 13, B: n = 13, C: n = 16)	7 (53.9)	7 (53.9)	8 (50.0)	0.97
Breastfeeding duration, months (Md, IQR) <sup>†</sup> (A: n = 13, B: n = 13, C: n = 16)	6 (3–10)	6 (3–12)	5.5 (1.5–7)	0.67
Day-care attendance	13 (86.7)	16 (94.1)	19 (79.2)	0.40
Shared bedroom <sup>†</sup> (B: n = 12, C: n = 21)	5 (33.3)	7 (58.3)	13 (61.9)	0.21
Parental smoking <sup>†</sup> (B: n = 15)	8 (53.3)	6 (40.0)	10 (41.7)	0.71
Vaccinated against influenza	0 (0.0)	0 (0.0)	1 (4.2)	1.00
Vaccinated against pneumococcal disease	9 (60.0)	13 (46.5)	14 (58.3)	0.45
PCV13 vaccinated	7 (77.8)	11 (84.6)	13 (92.9)	0.61
Respiratory infection in the previous 30 days <sup>†</sup> (n = 52)	6 (42.9)	5 (31.3)	15 (68.2)	0.07*
<b>MICROBIOLOGICAL VARIABLES</b>				
Sp detection in NP sample by PCR	7 (46.6)	9 (52.9)	18 (75.0)	0.16
High-attack rate Sp serotype in NP sample	3 (42.9)	3 (33.3)	6 (33.3)	0.90
PCV13 serotype coverage in NP sample	3 (42.9)	4 (44.4)	6 (33.3)	0.81
Viral detection in NP sample by PCR	7 (46.7)	8 (47.1)	16 (66.7)	0.34
Multiple respiratory viral infection	2 (28.6)	3 (37.5)	4 (25.0)	0.82
<b>IMMUNOLOGICAL VARIABLES (CASE GROUP)</b>				
White blood count, 10 <sup>3</sup> cells/mm <sup>3</sup> † (n = 21) (Md, IQR)	24.6 (18.9–33.4)	15.7 (10.0–18.2)	18.2 (7.5–22.7)	0.12
C-reactive protein, mg/L <sup>†</sup> (n = 20) (Md, IQR)	29.2 (10.8–69.0)	113.1 (50.5–237.6)	302.9 (261.5–355.8)	<b>0.009**</b>
<b>B</b>				
Variable	Microbiota A (n = 5)	Microbiota B (n = 14)	Microbiota C (n = 9)	p-value
<b>CLINICAL VARIABLES (CASE GROUP)</b>				
Underlying disease	1 (20.0)	1 (7.1)	1 (11.1)	0.75
Antibiotic intake in the previous 7 days <sup>†</sup> (n = 26)	5 (100.0)	12 (85.7)	6 (66.7)	1.00
Length of hospital stay, days (Md, IQR)	4 (3–6)	7 (7–8)	7.5 (3.5–10)	0.21
Admission to PICU	0 (0.0)	4 (28.6)	1 (11.1)	<b>0.53</b>
Length of PICU stay, days (Md, IQR)	0 (0.0)	2 (0–6)	1 (1–1)	0.45
Complications	0 (0.0)	2 (14.3)	2 (22.2)	0.80
<b>MAIN CLINICAL DIAGNOSIS</b>				
Complicated pneumonia	0 (0.0)	8 (57.1)	5 (55.6)	0.09
Bacteremia	3 (60.0)	3 (21.4)	1 (11.1)	0.17
Pneumonia	2 (40.0)	2 (14.3)	2 (22.2)	0.51
Meningitis	0 (0.0)	1 (7.1)	0 (0.0)	1.00
Sepsis	0 (0.0)	0 (0.0)	1 (11.1)	0.50
Serious IPD (complicated pneumonia, meningitis, or sepsis)	0 (0.0)	9 (64.9)	6 (66.7)	<b>0.04</b>
Course of disease, sequelae or exitus	0 (0.0)	3 (21.4)	2 (22.2)	0.67

Results expressed as counts (%), unless stated otherwise.

Md, median; IQR, Interquartile range; SD, standard deviation; NP, nasopharyngeal; PCV13, 13-valent pneumococcal conjugate vaccine; PICU, Pediatric Intensive Care Unit; IPD, Invasive Pneumococcal Disease.

<sup>†</sup>Missing values.

<sup>¶</sup>According to WHO recommendations.

\*p = 0.05 comparing Microbiota B and C.

\*\*p = 0.03 comparing Microbiota A and C; p = 0.04 comparing Microbiota A and B. The bold values means statistically significant associations with p-values < 0.05.

microbiotas (Table 3). However, the classification of patients into case and control groups was significantly associated with the nasopharyngeal-types ( $p = 0.006$ , Chi-square test). Among cases, nasopharyngeal-type B was the most frequently detected pattern (50.0%), followed by nasopharyngeal-type C (32.1%), while nasopharyngeal-type A was detected only in 17.9% of children with IPD. Conversely, nasopharyngeal-type B was only detected in 3 controls (10.7%).

Overall, children with nasopharyngeal-type A showed to have markedly lower inflammatory activity measured by the C-reactive protein level when compared to those with nasopharyngeal-type C ( $p = 0.03$ ) or nasopharyngeal-type B ( $p = 0.04$ ). In addition, no case with nasopharyngeal-type A was diagnosed with complicated pneumonia, meningitis, and sepsis, in contrast to the relatively high occurrence of these serious IPD manifestations among nasopharyngeal-type B and nasopharyngeal-type C cases (0.0, 64.9, and 66.7%, respectively,  $p = 0.04$ ). Similarly, none of the cases with nasopharyngeal-type A required intensive care but a noticeable proportion of cases with nasopharyngeal-type B and C did (0.0, 28.6, and 11.1%, respectively). Of note, 90% of the ten healthy controls with a Microbiota A were fed with maternal milk whereas breastfeeding was less frequent in controls with Microbiota B (33.3%) and C-types (60.0%). In spite of these differences, the association between breastfeeding and type of microbiota was not found to be statistically significant. On the other hand, there were no significant correlations between any specific bacterial genera over-represented in IPD patients and inflammatory parameters.

## DISCUSSION

The current work describes for the first time the nasopharyngeal microbiota in a case-control study of children with IPD and healthy children. Our data show that bacterial richness and diversity were significantly higher in IPD patients. In such cases, a clear dysbiosis was observed with a high frequency of *Veillonella* and other oral microorganisms which appeared to be relatively absent in controls. This over-representation of anaerobic and proteolytic oral species in children with IPD was also found by Segal et al. (2013) in adult bronchoalveolar lavage samples in association with increased inflammation. In our samples, the microbiota related with IPD was also associated with higher levels of the C-reactive protein inflammatory biomarker.

De Steenhuijsen Piters et al. (2016) similarly found higher nasopharyngeal microbiota diversity in elderly pneumonia patients compared to elderly healthy controls, whilst this difference was not found in adult patients. In ecological terms, these results are surprising since higher bacterial diversity is usually related with health (Turnbaugh et al., 2007), while lower bacterial diversity is associated with disease. It is noteworthy that patients with primary immunodeficiencies presented a higher microbiota diversity compared with healthy controls. This fact might correlate with an increase of immune system permissiveness for microbe colonization (Oh et al., 2013). Similarly, a local increase in bacterial diversity has been detected in polyps and tumor biopsies from patients with

colorectal cancer (Mira-Pascual et al., 2015), a disease which has been associated with immune suppression at the affected tissues. The best studied case of immune-driven dysbiosis is probably gum disease, where the presence of the “keystone pathogen” *Porphyromonas gingivalis* has been shown to induce a profound alteration of the immune system (Hajishengallis, 2011) facilitating the settlement of many species that produce inflammation. However, the increase in microbial diversity associated with dysbiosis in gum disease and cancer could not only be due to immune alteration but also to a nutritionally richer environment (Mira et al., 2017). In the respiratory tract of children with IPD, future investigations should elucidate whether the observed dysbiosis is a consequence of immune changes, increased nutrient availability or both. In periodontal disease, destructive inflammation generates abundant gum tissue breakdown products that serve as nutrients for proteolytic and saccharolytic bacteria to obtain essential amino-acids and iron, including degraded collagen, and heme-containing compounds (Hajishengallis et al., 2012). It is important to underline that *Porphyromonas* was one of the over-represented genera in IPD patients (Figure 2C). However, all reads belonging to this genus appeared to correspond to uncultured or unknown species (Table 2). We hope the present study stimulates research into the characterization of these potentially pathogenic species, including their potential involvement as keystone pathogens (Darveau et al., 2012).

In addition, IPD samples presented significantly higher levels of *Veillonella* (Figure 2C), which is a bacterium that uses lactate as a carbon source, as a consequence of which it is usually found physically and functionally associated to lactate producers like *Streptococcus* (Dige et al., 2014; Gaspar et al., 2014). In our data, however, the correlation between *Streptococcus* and *Veillonella* levels in cases was not significant ( $R^2 = 0.17$ ,  $p > 0.1$ , Supplementary Figure 3) and therefore it is unknown whether streptococcal serotypes associated with IPD produce more lactate than those found in healthy children. Future work should determine whether the seemingly higher acidic environment in children with IPD is a consequence of lactate production by microorganisms, by lactate-dehydrogenase over-expression in the human tissue (Tan et al., 2002) or both.

Our data identified three bacterial clusters or nasopharyngeal-types, dominated by *Dolosigranulum*, *Streptococcus*, or *Haemophilus*. Similarly, other studies have described the presence of bacterial nasopharyngeal clusters. In a Danish cohort (1–3 months of age) and a Dutch cohort (from birth to 6 months of age), up to 5 and 4 nasopharyngeal-types were described (Mortensen et al., 2016), respectively. Although there are methodological, geographical and age-related differences between these studies, all available data from respiratory tract samples point toward the presence of precise microbial communities with a given dominant bacteria on each nasopharyngeal-type, although larger samples sizes are required to validate this.

In the current manuscript, the only statistically significant explanatory factor for these three microbiota profiles was the classification of patients into case or control groups. *Streptococcus*-dominated community (Microbiota B) was clearly

associated to IPD patients. Given that *Dolosigranulum* abundance in the nasopharynx has been reported to be inversely associated with episodes of wheezing and mild respiratory tract infections (Biesbroek et al., 2014) we speculate that the *Dolosigranulum*-dominated community (Microbiota A) could be more resistant to pneumococcal infection occurrence and severity. Our hypothesis was supported by the lower number of cases with a Microbiota A-type (17.9% of the total) and by the fact that such cases experienced less severe manifestations of IPD and did not require intensive care. In this respect, it is suggestive that *Dolosigranulum* and *Haemophilus* are found in the present study to have a negative correlation with *Streptococcus*, the dominant genera in IPD cases (Figure 4). Moreover, the nasopharyngeal presence of *Dolosigranulum* has been associated with breastfeeding. It has been shown that lactation has a profound impact in the microbial community composition of the infant upper respiratory tract, increasing the prevalence and levels of *Dolosigranulum* and reducing the levels of *Staphylococcus*, *Prevotella*, or *Veillonella* (Biesbroek et al., 2014). In agreement with this, we found a considerably higher proportion of Microbiota A healthy controls that were breastfed and a trend for statistical significance in the relation between exclusive breastfeeding up to 6 months of age and case-control classification ( $p = 0.06$ ) (Table 1). Despite the lack of statistical power to observe significance in these associations, the results obtained suggest that breastfeeding could have an important role in protection against pneumococcal infection. As far as for *Haemophilus*-dominated community (microbiota C), both IPD cases and healthy controls appear to fall within this microbiota profile. This genus, together with *Moraxella*, has been associated with higher rates of parent-reported upper respiratory infections and wheezing in the first years of life (Hyde et al., 2014; Bogaert et al., 2015). In fact, the group of children with *Haemophilus* and *Moraxella*-dominated microbiota showed a higher percentage of patients presenting a respiratory infection in the previous 30 days before the sample collection (68.2%), with a trend for significance ( $p = 0.07$ ). However, future experimental work with a larger sample size should further explore the significance of these microbiota profiles and test whether the presence of a given nasopharyngeal microbiota makes individuals more sensitive or resistant to pneumococcal infection.

Our study describes the use of a long-amplicon sequencing approach of the 16S rRNA for species-level identification. Computer simulations show that taxonomic assignment accuracy, especially at the species level, decreases dramatically in short reads, like those of current Illumina or Ion Torrent technologies (Claesson et al., 2010), while the use of long-amplicon ( $\geq 500$  bp) has been postulated as a good tool for species-level taxonomical assignments (Shin et al., 2016) and has been widely used as a diagnostic tool of bacterial infections in clinical samples (Guembe et al., 2013; Martinez et al., 2016). However, we have found surprising results using 16S rRNA for species level identification since *S. pneumoniae* would be more frequent in controls than in cases. These results are discordant with the results generated from a real-time PCR targeting *LytA* gene, the target recommended by CDC for pneumococcal detection. *LytA* PCR indicated the presence of *S. pneumoniae* in 64.4% of cases vs. 57.1% in controls (Table 1). According to

our results, taxonomic assignment by using 16S rRNA could be limited in some genera and accurate identification at the species level should be done with more conserved genes. It is important to keep in mind that streptococci are particularly similar in the 16S rRNA gene sequence, and therefore species-level assignment in this genus must be treated with caution (Ing et al., 2012). Even more, despite *LytA* is the recommended gene utilized for pneumococcal diagnosis according to CDC (National Center for Immunization and Respiratory Diseases, 2014), this target has been reported to misidentify the pathogen (Morales et al., 2015; Simões et al., 2016), may be showing the occurrence of genetic exchange among streptococcal species and opening the possibility that other etiological agents could be involved. Thus, future work should increase our understanding of these microbial consortia by accurate species-level identification.

## CONCLUSION

In conclusion, we found a higher bacterial diversity and richness in children with IPD which could suggest an impaired immune response. This lack of immune competence could be aggravated by limited breastfeeding lower than recommended by WHO and by the presence of keystone pathogens which need to be characterized and that could favor the overgrowth of many proteolytic anaerobic organisms from the oral cavity, giving rise to a dramatic dysbiosis. From an applied point of view, we found suggestive microbiota profiles associated to IPD (*Streptococcus*-dominated microbiota profile) or asymptomatic colonization (*Dolosigranulum*-dominated microbiota profile) that could be used, respectively, as disease biomarkers or to identify health-associated inhabitants of the respiratory tract. The characterization of beneficial bacteria could be useful to prevent pneumococcal infections by integrating those microorganisms in a probiotic formula. The present study suggests not only respiratory tract samples, but also breast milk, as a potential source of those beneficial bacteria.

## DATA AVAILABILITY STATEMENT

Sequence files and metadata for all samples used in this study are stored in the MG-RAST server to be publicly available by accessing the project Invasive Pneumococcal Disease (IPD) in children is associated with a highly diverse nasopharyngeal microbiota: a case-control study, ID mgp80930. All remaining data are contained within the paper and/or Supporting Information Files.

## AUTHOR CONTRIBUTIONS

AM and CM-A contributed to the conception, design, and analysis of data. DH and JR contributed to the case and control recruitment and microbiological characterization of samples. AC-C and AG contributed to the bioinformatics analysis. PB contributed to the coordination, statistical analysis, and data/sample management. AC-C, PB, AM, and CM-A were the major contributors in writing the manuscript. All authors read and approved the final manuscript.

## FUNDING

This work was supported in part by Fondo Europeo de Desarrollo Regional (FEDER) and the Ministry of Science and Innovation, Instituto de Salud Carlos III (ISCIII) (<http://www.isciii.es>), Projects of Research on Health [FIS project number PI16/00174, Principal Investigator: CM-A], Predoctoral contract for training in research into Health [Grant number FI17/00248, Recipient: DH] and Sant Joan de Deu Foundation [Grant number AFR2015, Principal Investigator: CM-A] (<http://www.fsjd.org/es>). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## ACKNOWLEDGMENTS

We thank A. Artacho for his help in statistical analysis and The Catalan Study group of host-pathogen interaction in patients with IPD: Maria Andres-Franch, Carlos Bautista, PB, AC-C, Alvaro Díaz, Cristina Esteva, Mariona F. de Sevilla, Ana Fernandez, Miguel Garcia-Fernandez, AG, Carmen Galles, Lourdes Garcia, Juan Jose Garcia-Garcia, Susanna

Hernandez-Bou, Cristian Launes, Francisco Lozano, Carles Luaces, AM, CM-A, Marian Navarro, Daniel Penela, JR, Miriam Triviño, Ana Valero-Rello, Goretti Sauca, Eudal Sellares.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2019.00011/full#supplementary-material>

**Supplementary Figure 1** | Rarefaction curves of individual nasopharyngeal samples in (a) healthy children (controls) and (b) patients with invasive pneumococcal disease (cases). The horizontal axis shows the number of reads (sequencing effort) obtained by pyrosequencing the 16S rRNA gene. The vertical axis shows the number of operational taxonomic units (OTUs) at a level of 97% (estimated number of bacterial species per sample).

**Supplementary Figure 2** | Heatmap profile showing average abundances of bacterial genera in cases and controls indicating the age of each child. Genera present in <3 patients were removed from the analysis for clarity.

**Supplementary Figure 3** | Relationships between the levels of *Streptococcus* and *Veillonella* in children's nasopharyngeal samples. Scatterplots show the number of sequence reads in the two bacterial genera for cases (A) and controls (B).

## REFERENCES

- Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., et al. (2011). Enterotypes of the human gut microbiome. *Nature* 473, 174–180. doi: 10.1038/nature09944
- Ballard, O., and Morrow, A. L. (2013). Human milk composition. *Pediatr. Clin. North Am.* 60, 49–74. doi: 10.1016/j.pcl.2012.10.002
- Biesbroek, G., Bosch, A. A. T. M., Wang, X., Keijsers, B. J. F., Veenhoven, R. H., Sanders, E. A., et al. (2014). The Impact of breastfeeding on nasopharyngeal microbial communities in infants. *Am. J. Respir. Crit. Care Med.* 190, 298–308. doi: 10.1164/rccm.201401-0073OC
- Bogaert, D., De Steenhuijsen Piters, W. A. A., and Sanders, E. A. M. (2015). The role of the local microbial ecosystem in respiratory health and disease. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140294. doi: 10.1098/rstb.2014.0294
- Bogaert, D., Keijsers, B., Huse, S., Rossen, J., Veenhoven, R., van Gils, E., et al. (2011). Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. *PLoS ONE* 6:e17035. doi: 10.1371/journal.pone.0017035
- Bomar, L., Brugger, S. D., Yost, B. H., Davies, S. S., and Lemon, K. P. (2016). *Corynebacterium accolens* releases antipneumococcal free fatty acids from human nostril and skin surface triacylglycerols. *mBio* 7, e01725–e01715. doi: 10.1128/mBio.01725-15
- Brotans, P., Bassat, Q., Lanaspá, M., Henares, D., Perez-Arguello, A., Madrid, L., et al. (2017). Nasopharyngeal bacterial load as a marker for rapid and easy diagnosis of invasive pneumococcal disease in children from Mozambique. *PLoS ONE* 12:e0184762. doi: 10.1371/journal.pone.0184762
- Brueggemann, A. B., Griffiths, D. T., Meats, E., Peto, T., Crook, D. W., and Spratt, B. G. (2003). Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential. *J. Infect. Dis.* 187, 1424–1432. doi: 10.1086/374624
- Cabrera-Rubio, R., Collado, M. C., Laitinen, K., Salminen, S., Isolauri, E., and Mira, A. (2012). The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. *Am. J. Clin. Nutr.* 96, 544–551. doi: 10.3945/ajcn.112.037382
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scand. J. Stat.* 11, 265–270.
- Claesson, M. J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J. R., Ross, R. P., et al. (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* 38:e200. doi: 10.1093/nar/gkq873
- Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., et al. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37, D141–D145. doi: 10.1093/nar/gkn879
- Darveau, R. P., Hajishengallis, G., and Curtis, M. A. (2012). Porphyromonas gingivalis as a potential community activist for disease. *J. Dent. Res.* 91, 816–820. doi: 10.1177/0022034512453589
- de la Cuesta-Zuluaga, J., and Escobar, J. S. (2016). Considerations for optimizing microbiome analysis using a marker gene. *Front. Nutr.* 3:26. doi: 10.3389/fnut.2016.00026
- De Steenhuijsen Piters, W. A. A., Huijskens, E. G. W., Wyllie, A. L., Biesbroek, G., van den Bergh, M. R., Veenhoven, R. H., et al. (2016). Dysbiosis of upper respiratory tract microbiota in elderly pneumonia patients. *ISME J.* 10, 97–108. doi: 10.1038/ismej.2015.99
- del Amo, E., Brotans, P., Monsonis, M., Triviño, M., Iñigo, M., Selva, L., et al. (2014). High invasiveness of pneumococcal serotypes included in the new generation of conjugate vaccines. *Clin. Microbiol. Infect.* 20, 684–689. doi: 10.1111/1469-0691.12422
- del Amo, E., Selva, L., de Sevilla, M. F., Ciruela, P., Brotans, P., Triviño, M., et al. (2015). Estimation of the invasive disease potential of *Streptococcus pneumoniae* in children by the use of direct capsular typing in clinical specimens. *Eur. J. Clin. Microbiol. Infect. Dis.* 34, 705–711. doi: 10.1007/s10096-014-2280-y
- Dethlefsen, L., McFall-Ngai, M., and Relman, D. A. (2007). An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* 449, 811–818. doi: 10.1038/nature06245
- Di Pierro, F., Colombo, M., Zanvit, A., Risso, P., and Rottoli, A. S. (2014). Use of *Streptococcus salivarius* K12 in the prevention of streptococcal and viral pharyngotonsillitis in children. *Drug. Healthc. Patient Saf.* 6, 15–20. doi: 10.2147/DHPS.S59665
- Dige, I., Grønkvær, L., and Nyvad, B. (2014). Molecular studies of the structural ecology of natural occlusal caries. *Caries Res.* 48, 451–460. doi: 10.1159/000357920
- Dray, S., and Dufour, A.-B. (2007). *The ade4 Package: Implementing the Duality Diagram for Ecologists*. Available online at: <http://www.jstatsoft.org/> (Accessed January 14, 2019).

- Gaspar, P., Al-Bayati, F. A. Y., Andrew, P. W., Neves, A. R., and Yesilkaya, H. (2014). Lactate dehydrogenase is the key enzyme for pneumococcal pyruvate metabolism and pneumococcal survival in blood. *Infect. Immun.* 82, 5099–5109. doi: 10.1128/IAI.02005-14
- Guembe, M., Marin, M., Martin-Rabadan, P., Echenagusia, A., Camunez, F., Rodriguez-Rosales, G., et al. (2013). Use of universal 16S rRNA gene PCR as a diagnostic tool for venous access port-related bloodstream infections. *J. Clin. Microbiol.* 51, 799–804. doi: 10.1128/JCM.02414-12
- Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., et al. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* 21, 494–504. doi: 10.1101/gr.112730.110
- Hajishengallis, G. (2011). Immune evasion strategies of *Porphyromonas gingivalis*. *J. Oral Biosci.* 53, 233–240. doi: 10.2330/joralbiosci.53.233
- Hajishengallis, G., Darveau, R. P., and Curtis, M. A. (2012). The keystone-pathogen hypothesis. *Nat. Rev. Microbiol.* 10, 717–725. doi: 10.1038/nrmicro2873
- Hunt, K. M., Foster, J. A., Forney, L. J., Schü Tte, U. M. E., Beck, D. L., Abdo, Z., et al. (2011). Characterization of the diversity and temporal stability of bacterial communities in human milk. *PLoS ONE* 6:e21313. doi: 10.1371/journal.pone.0021313
- Hyde, E. R., Petrosino, J. F., Piedra, P. A., Camargo, C. A., Espinola, J. A., Mansbach, J. M., et al. (2014). Nasopharyngeal Proteobacteria are associated with viral etiology and acute wheezing in children with severe bronchiolitis. *J. Allergy Clin. Immunol.* 133, 1220–1222. doi: 10.1016/j.jaci.2013.10.049
- Ing, J., Mason, E. O., Kaplan, S. L., Lamberth, L. B., Revell, P. A., Luna, R. A., et al. (2012). Characterization of nontypeable and atypical *Streptococcus pneumoniae* pediatric isolates from 1994 to 2010. *J. Clin. Microbiol.* 50, 1326–1330. doi: 10.1128/JCM.05182-11
- Jeffery, I. B., Claesson, M. J., Toole, P. W. O., and Shanahan, F. (2012). Categorization of the gut microbiota: enterotypes or gradients? *Nat. Rev. Microbiol.* 10, 591–592. doi: 10.1038/nrmicro2859
- López-López, A., Camelo-Castillo, A., Ferrer, M. D., Simon-Soro, Á., and Mira, A. (2017). Health-associated niche inhabitants as oral probiotics: the case of *Streptococcus dentisani*. *Front. Microbiol.* 8:379. doi: 10.3389/fmicb.2017.00379
- Lozupone, C., Hamady, M., and Knight, R. (2006). UniFrac—an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* 7:371. doi: 10.1186/1471-2105-7-371
- Martínez, M. J., Massora, S., Mandomando, I., Ussene, E., Jordao, D., Lovane, L., et al. (2016). Infectious cause of death determination using minimally invasive autopsies in developing countries. *Diagn. Microbiol. Infect. Dis.* 84, 80–86. doi: 10.1016/j.diagmicrobio.2015.10.002
- Medina, M., Villena, J., Salva, S., Vintiñi, E., Langella, P., and Alvarez, S. (2008). Nasal administration of *Lactococcus lactis* improves local and systemic immune responses against *Streptococcus pneumoniae*. *Microbiol. Immunol.* 52, 399–409. doi: 10.1111/j.1348-0421.2008.00050.x
- Meningitis (2011). *Laboratory Methods for the Diagnosis of Meningitis*. CDC. Available online at: <https://www.cdc.gov/meningitis/lab-manual/index.html> (Accessed January 14, 2019).
- Michener, C. D., and Sokal, R. R. (1957). A quantitative approach to a problem in classification. *Evolution* 11, 130–162. doi: 10.2307/2406046
- Mira, A., Simon-Soro, A., and Curtis, M. A. (2017). Role of microbial communities in the pathogenesis of periodontal diseases and caries. *J. Clin. Periodontol.* 44, S23–S38. doi: 10.1111/jcpe.12671
- Mira-Pascual, L., Cabrera-Rubio, R., Ocon, S., Costales, P., Parra, A., Suarez, A., et al. (2015). Microbial mucosal colonic shifts associated with the development of colorectal cancer reveal the presence of different bacterial and archaeal biomarkers. *J. Gastroenterol.* 50, 167–179. doi: 10.1007/s00535-014-0963-x
- Morales, M., Martín-Galiano, A. J., Domenech, M., and García, E. (2015). Insights into the evolutionary relationships of LytA autolysin and ply pneumolysin-like genes in *Streptococcus pneumoniae* and related Streptococci. *Genome Biol. Evol.* 7, 2747–2761. doi: 10.1093/gbe/evv178
- Mortensen, M. S., Brejnrod, A. D., Roggenbuck, M., Abu Al-Soud, W., Balle, C., Krogfelt, K. A., et al. (2016). The developing hypopharyngeal microbiota in early life. *Microbiome* 4:70. doi: 10.1186/s40168-016-0215-9
- National Center for Immunization and Respiratory Diseases (2014). *PCR for Detection and Characterization of Bacterial Meningitis Pathogens: Neisseria meningitidis, Haemophilus influenzae, and Streptococcus pneumoniae*. Oh, J., Freeman, A. F., NISC Comparative Sequencing Program., Park, M., Sokolic, R., Candotti, F., et al. (2013). The altered landscape of the human skin microbiome in patients with primary immunodeficiencies. *Genome Res.* 23, 2103–2114. doi: 10.1101/gr.159467.113
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2018). *Community Ecology Package*. Available at: <https://cran.r-project.org/web/packages/vegan/vegan.pdf> (Accessed November 22, 2018).
- Relman, D. A. (2012). The human microbiome: ecosystem resilience and health. *Nutr. Rev.* 70, S2–S9. doi: 10.1111/j.1753-4887.2012.00489.x
- Segal, L. N., Alekseyenko, A. V., Clemente, J. C., Kulkarni, R., Wu, B., Gao, Z., et al. (2013). Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome* 1:19. doi: 10.1186/2049-2618-1-19
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12:R60. doi: 10.1186/gb-2011-12-6-r60
- Selva, L., Berger, C., Garcia-Garcia, J. J., de Paz, H., Nadal, D., and Muñoz-Almagro, C. (2014). Direct identification of *Streptococcus pneumoniae* capsular types in pleural fluids by using multiplex PCR combined with automated fluorescence-based capillary electrophoresis. *J. Clin. Microbiol.* 52, 2736–2737. doi: 10.1128/JCM.00906-14
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423. doi: 10.1145/584091.584093
- Shin, J., Lee, S., Go, M. J., Lee, S. Y., Kim, S. C., Lee, C. H., et al. (2016). Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing. *Sci. Rep.* 6, 1–10. doi: 10.1038/srep29681
- Simões, A. S., Tavares, D. A., Rolo, D., Ardanuy, C., Goossens, H., Henriques-Normark, B., et al. (2016). LytA-based identification methods can misidentify *Streptococcus pneumoniae*. *Diagn. Microbiol. Infect. Dis.* 85, 141–148. doi: 10.1016/j.diagmicrobio.2016.03.018
- Simón-Soro, A., Guillen-Navarro, M., and Mira, A. (2014). Metatranscriptomics reveals overall active bacterial composition in caries lesions. *J. Oral Microbiol.* 6:25443. doi: 10.3402/jom.v6.25443
- Sleeman, K. L., Griffiths, D., Shackley, F., Diggle, L., Gupta, S., Maiden, M. C., et al. (2006). Capsular serotype-specific attack rates and duration of carriage of *Streptococcus pneumoniae* in a population of children. *J. Infect. Dis.* 194, 682–688. doi: 10.1086/505710
- Tagg, J. R., and Dierksen, K. P. (2003). Bacterial replacement therapy: adapting “germ warfare” to infection prevention. *Trends Biotechnol.* 21, 217–223. doi: 10.1016/S0167-7799(03)00085-4
- Tan, T. Q., Mason, E. O., Wald, E. R., Barson, W. J., Schutze, G. E., Bradley, J. S., et al. (2002). Clinical characteristics of children with complicated pneumonia caused by *Streptococcus pneumoniae*. *Pediatrics* 110, 1–6. doi: 10.1542/peds.110.1.1
- Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The human microbiome project. *Nature* 449, 804–810. doi: 10.1038/nature06244
- Yarza, P., Richter, M., Peplies, J., Euzéby, J., Amann, R., Schleifer, K.-H., et al. (2008). The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst. Appl. Microbiol.* 31, 241–250. doi: 10.1016/j.syapm.2008.07.001
- Zhu, B., Wang, X., and Li, L. (2010). Human gut microbiome: the second genome of human body. *Protein Cell* 1, 718–725. doi: 10.1007/s13238-010-0093-z

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Camelo-Castillo, Henares, Brotons, Galiana, Rodriguez, Mira and Muñoz-Almagro. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.