Deep 16S rRNA metagenomics and quantitative PCR analyses of the premature infant fecal microbiota

Running title: Intestinal microbiota in premature infants

Silvia Arboleya[1], Li Ang[2], Abelardo Margolles[1], Li Yiyuan[2], Zhang Dongya[2], Xiao Liang[2], Gonzalo Solís [3], Nuria Fernández [4], Clara G. de los Reyes-Gavilán[1], and Miguel Gueimonde[1*].

[1] Department of Microbiology and Biochemistry of Dairy Products, Instituto de Productos Lácteos de Asturias (IPLA-CSIC), Villaviciosa, Asturias, Spain.

[2] Beijing Genomics Institute. Beishan Industrial Zone, Yantian District, Shenzhen 518083, China.

[3] Pediatrics Service, Hospital Universitario Central de Asturias, SESPA, Oviedo, Asturias, Spain

[4] Pediatrics Service, Hospital de Cabueñes, SESPA, Gijón, Asturias, Spain.

* Corresponding author: Miguel Gueimonde. Department of Microbiology and Biochemistry of Dairy Products. Instituto de Productos Lácteos de Asturias. Consejo Superior de Investigaciones Científicas (IPLA-CSIC). Ctra. Infiesto s/n, 33300 Villaviciosa, Asturias, Spain. Tel. +34 985892131, Fax. +34 985892233. E-mail; mgueimonde@ipla.csic.es

All authors disclose any conflict of interest

Abstract

Metagenomic studies on the gut microbiota of preterm infants are scarce. We characterized the microbiota of 10 days-old neonates by deep 16S rRNA gene metagenomic analysis and compare the results with those obtained by qPCR. Both techniques lead to similar conclusions, allowing differentiating between preterm and full-term infants.

Key words: *Intestinal microbiota, preterm infants, neonates, metagenomics*

The human gastrointestinal tract harbors a very complex microbial community. Despite the high inter-individual variability, recent metagenomic studies have identified different microbiota enterotypes in humans [1], and microbiota alterations related with diseases such as obesity [2]. However, most metagenomic studies have focused on adult populations, the reports in neonates being still scarce. Microbial colonization of the digestive tract starts at birth and provides essential signals for intestinal development and immune maturation [3] and, therefore, the step of establishment of this microbiota may have a profound effect on later health. The first metagenomic studies on the development of infant intestinal microbiota dealt with full-term infants [4,5] and only recently some data, mainly focusing on the relationship between microbiota and necrotizing enterocolitis, became available on preterm-infants [6,7].

Moreover, in spite of the increasing use of metagenomic techniques it is not known to which extent the results are comparable with those previously obtained by other widely used molecular methods. To this regard, some studies have compared pyrosequencing with phylogenetic microarrays finding similar fecal microbiota profiles [8,9]; however studies comparing metagenomics with other techniques such as quantitative PCR (qPCR) are still lacking.

In the present work we aimed at characterizing the microbiota composition in 10 days-old infants by deep 16S rRNA gene metagenomic analysis, and to compare the results with those obtained with the same fecal samples by qPCR.

We collected fecal samples from 6 vaginally delivered babies at 10 days of age. Two of them were very-low-birth-weight (VLBW) preterm babies; 1 female (VLBW1) and 1 male (VLBW2) with gestational ages 27 and 28 weeks (discharged from the hospital after 62 and 64 days) and birth weights 1050 and

1315 gr, respectively. Both infants received ampicillin plus gentamicin from birth to 6 and 8 days of life, respectively, and were on mixed feeding (received both breast-milk and formula). In addition, the mother of the last infant (VLBW2) received intrapartum ampicillin and gentamicin. Two babies were born prematurely (P) (1 female [P1]/1 male [P2]; gestational ages 32/34 weeks; birth weights 2115/1820 gr), they were on mixed feeding and were discharged after 20 and 23 days of hospitalization, respectively. Mothers of both babies received intrapartum ampicillin, whilst none of the babies received antibiotics. Finally two infants were full-term (F) (1 female [F1]/1 male [F2]; gestational age 39 weeks; birth weights 3325/3130 gr), they were exclusively breast-fed, discharged from the hospital at the third day of life, and were not exposed to antibiotics. The study was approved by the Regional Ethical Committee of Asturias Public Health Service (SESPA) and written informed consent was obtained from each mother.

Fecal samples were aseptically taken and immediately frozen until analyses. Fecal DNA was extracted as previously indicated [10] and DNA was kept at -80ºC until analyses. 5 ng of DNA were used for PCRs. Amplification of the V6 region of the 16S rRNA gene was obtained by using universal primers described elsewhere [11] whilst group-specific qPCRs (*Enterobacteriaceae*, *Enterococcaceae*, *Streptococcaceae*, *Staphylococcaceae*, *Bacteroidaceae*, *Bifidobacteriaceae*, *Lactobacillaceae*, *Clostridium leptum*-group and *Clostridium coccoides*-group) were carried out as previously described [10]. For metagenomic analyses, in order to avoid PCR biases, the V6 region PCR consisted of only 20 amplification cycles. PCR products were purified by using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). 5 µg of amplified

16S rDNA were submitted for sequencing in an Illumina genome analyzer Hiseq 2000 (101 bp pair-end sequencing strategy). Altogether 11.9 GB of raw data were obtained. The base-calling pipeline (version Solexapipeline-0.3) was used to process the raw fluorescent images and sequences.

Sequences with more than 3 mismatches within amplification primers' regions were removed, then we trimmed bases with the lowest quality value (bases with 0.63 sequencing error rate) in the 3'end of each sequence. Trimmed sequences were overlapped to the V6 tags. The minimum overlap length was 30 bp. Finally, over 6.5 million tags were obtained from each sample (average 7.972.938 per sample), rendering over 28000 unique tags per sample. Tags were then classified by aligning them to Silva RefSSU database [12] using BLASTN, with the criterion of "-p blastn –F F –e 1e-5", and selecting the best alignments. Sample tags that did not have a BLAST match in the Silva RefSSU database were not given a taxonomic assignment. We used a two-thirds (66%) majority rule [13], meaning that when more than 66% of alignments belonged to a certain taxonomic group, the tag was classified as that group, and then the analysis went into the next taxonomic rank. The abundance of tags in different classification levels was calculated and according to the results, we choose family level as the optimal level for further analyses. For estimating the diversity in the environment, we performed an Operational Taxonomic Unit (OTU) analysis using mothur program package (version 1.14.0) [14]. We got the OTUs at 97% identity obtaining on average 12444, 14819 and 25317 clusters for VLBW, preterm and full-term infants, respectively. A higher α-diversity was observed for full-term (Shannon and Simpson indexes; average 3.287 and

0.123, respectively) than for preterm (1.901 and 0.429) or VLBW preterm infants (1.312 and 0.584, respectively).

At family level *Enterobacteriaceae* was by far the predominant group in preterm babies, either normal weight or VLBW, whilst in full-term infants other microorganisms such as *Streptococcaceae*, *Lactobacillaceae* or *Bifidobacteriaceae* were also among the dominating populations (Figure 1, supplementary file 1).

In order to compare the results achieved by metagenomics with those obtained by qPCR the values for the different microbial groups analyzed by both techniques were relativized to the total sum of sequences, for metagenomics, and to that of bacteria per gram, for qPCR results. According to the metagenomic results the groups included in this analysis represented, on average, a 95.21% of the total fecal microbiota in the 6 samples analyzed (sum of sequences belonging to these groups with regard to the total number of sequences). The results were subjected to principal component analyses (PCA) using function “princomp” in R program package. Both techniques, allowed a clear distinction of preterm from full-term infants (Figure 2A), although metagenomics rendered lower bifidobacterial levels than qPCR (especially in full-term infants, data not shown). An underestimation of bifidobacteria when using metagenomics has been previously described [15]. When the different samples, analyzed by both methods, were submitted to cluster analysis (“hclust” function in R program, using distances of “Hellinger” and “ward” cluster method) (Figure 2B) the general agreement between both techniques was confirmed, with the samples from full-term infants clustering together, and far from those of

preterm neonates, independently on the technique used. Therefore, both techniques led to a similar biological conclusion.

Although the number of infants analyzed and the presence of potential confounding factors (i.e. antibiotic use, feeding habits or hospital stay) do not allow establishing firm conclusions our results appear to confirm both, the dominance of *Enterobacteriaceae* in the fecal microbiota of preterm infants, as previous metagenomic studies have indicated [6], and a reduced microbial diversity in these infants [10]. Microbiota from VLBW preterm infants did not show major differences with regard to that of non-VLBW premature babies, whilst both groups of infants were clearly different from full-term infants. In addition, our results show that both techniques used, metagenomics and qPCR of the main intestinal microbial groups, provided similar results, leading to the same allocation of the samples and allowing distinguishing between preterm and full-term infants.

This work was funded by a CSIC (Spain) intramural project (Ref. 200870I049). S. A. was funded by a predoctoral JAE fellowship from CSIC, Spain.

References

[1] M. Arumugam, J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D.R. Mende, et al., Enterotypes of the human gut microbiome, Nature 473 (2011) 174-180.

[2] P.J. Turnbaugh, M. Hamady, T. Yatsunenko, B.L. Cantarel, A. Duncan, R.E. Ley, et al., A core gut microbiome in obese and lean twins, Nature 457 (2009) 480-484.

[3] L.V. Hooper, A.J. Macpherson, Immune adaptations that maintain homeostasis with the intestinal microbiota, Nat. Rev. Immunol. 10 (2010) 159-169.

[4] J.E. Koenig, A. Spor, N. Scalfone, A.D. Fricker, J. Stombaugh, R. Knight, et al., Succession of microbial consortia in the developing infant gut microbiome, Proc. Natl. Acad. Sci. USA 108 (2011) 4578-4585.

[5] C. Palmer, E.M. Bik, D.B. DiGiulio, D.A. Relman, P.O. Brown, Development of the human infant intestinal microbiota, PLoS Biol. 5 (2007) e177.

[6] W. Mai, C.M. Young, M. Ukhanova, X. Wang, Y. Sun, G. Casella, et al., Fecal microbiota in premature infants prior to necrotizing enterocolitis, PLoS One 6 (2011) e20647.

[7] M.J. Morowitz, V.J.Denet, E.K. Costello, B.C. Thomas, V. Poroyko, D.A. Relman, J.F. Banfield, Strain-resolved community genomic analysis of gut microbial colonization in a premature infant, Proc. Natl. Acad. Sci. USA 108 (2011) 1128-1133.

[8] M. Claesson, O. O'Sullivan, Q. Wang, J. Nikkila, J.R. Marchesi, H. Smidt, et al., Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine, PLoS Immunol. 4 (2009) e6669.

[9] B. Van der Bogert, W.M. de Vos, E.G. Zoetendal, M. Kleerebezem, Microarray analysis and barcoded pyrosequencing provide consistent microbial profiles depending on the source of human intestinal samples, Appl. Environ. Microbiol. 77 (2011) 2071-2080.

[10] S. Arboleya, A. Binetti, N. Salazar, N. Fernández, G. Solís, A. Hernandez-Barranco, et al., Establishment and development of intestinal microbiota in preterm neonates, FEMS Microbiol. Ecol. 79 (2012) 763-772.

[11] M.L. Sogin, H.G. Morrison, J.A. Huber, D.A. Welch, S.M. Huse, P.R. Neal, et al., Microbial diversity in the deep sea and the underexplored “rare biosphere”, Proc. Natl. Acad. Sci. USA 103 (2006) 12115-12120.

[12] E. Pruesse, C. Quast, K. Knittel, B. Fuchs, W. Ludwig, J. Peplies, F.O. Glöckner, SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB, Nucl. Acids Res. 35 (2007) 7188-7196.

[13] S.M. Huse, L. Dethlefsen, J.A. Huber, D M. Welch, D.A. Relman, M.L. Sogin, Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing, PLoS Genet. 4 (2008) e1000255.

[14] P.D. Schloss, S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, et al., Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities, Appl. Environ. Microbiol. 75 (2009) 7537-7541.

[15] Z.A. Khachatryan, Z.A. Ktsoyan, G.P. Manukyan, D. Kelly, K.A. Ghazaryan, R.I. Aminov, Predominant role of host genetics in controlling the composition of gut microbiota, PLoS One 3 (2008) e3064.

Figure legends

Figure 1. Percentage (%) of the most abundant bacterial families in the six fecal samples analyzed. VLBW, very-low-birth-weight premature infant; P, premature infant; F, full-term infant.

Figure 2. (A) Principal Components Analysis (PCA) and (B) dendrogram obtained by cluster analysis of the six fecal samples analyzed by either metagenomics or qPCR. VLBW, very-low-birth-weight premature infant; P, premature infant; F, full-term infant.

Supplementary file 1. Number of tags, at family level, obtained for the 6 fecal samples analyzed by 16S rDNA metagenomics.

Figure1

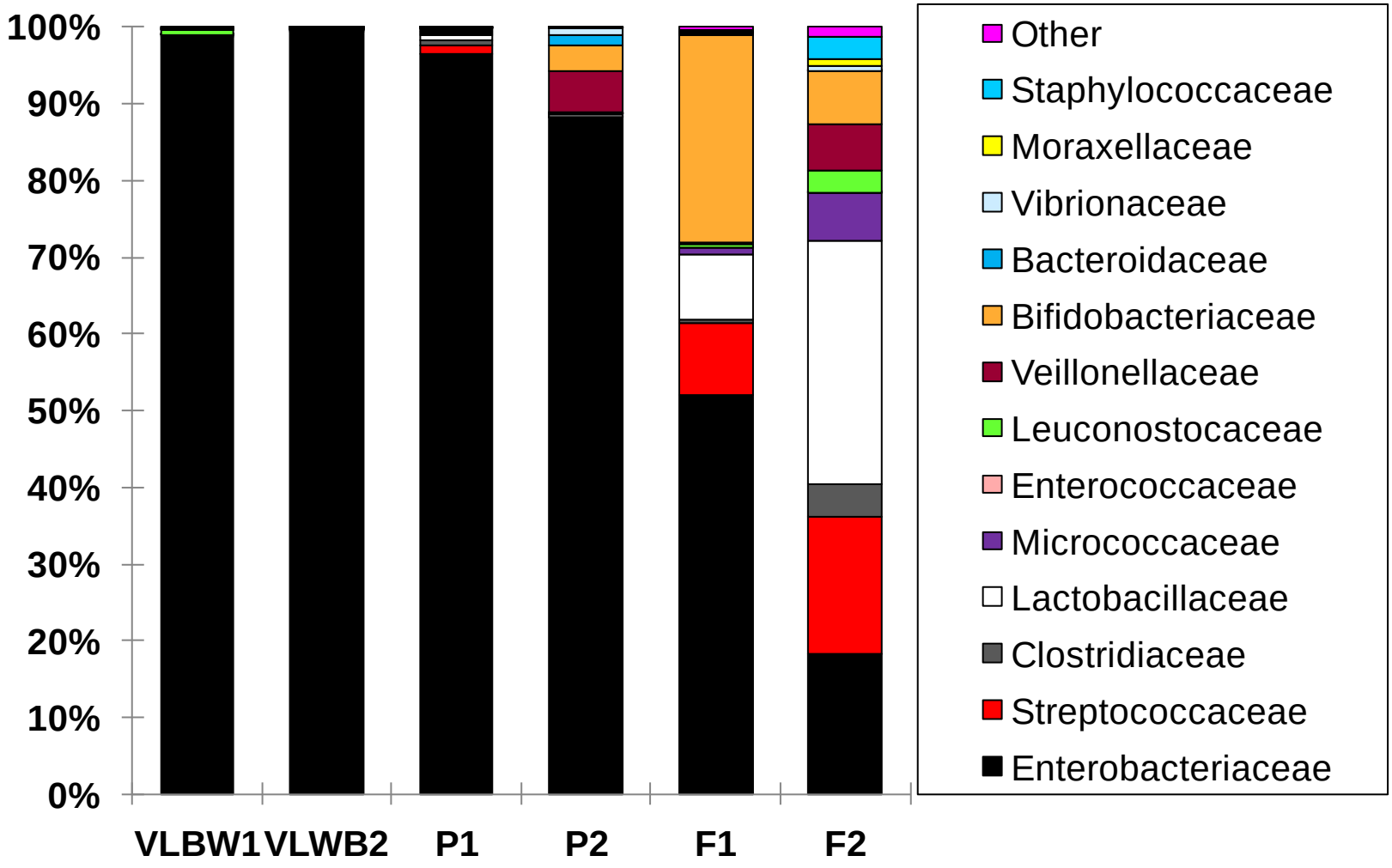
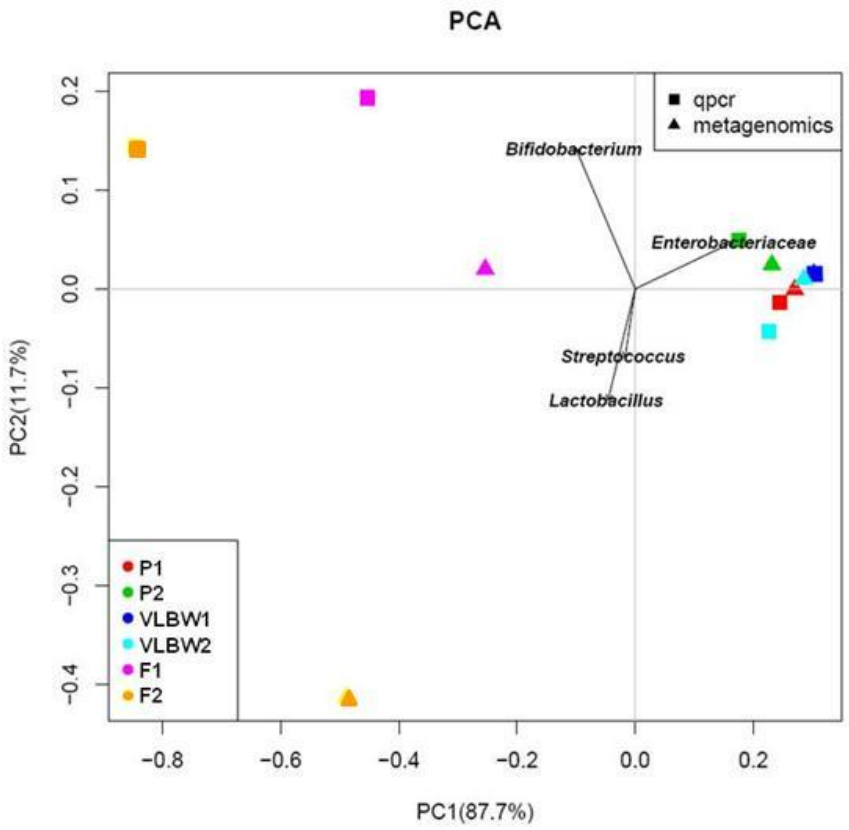
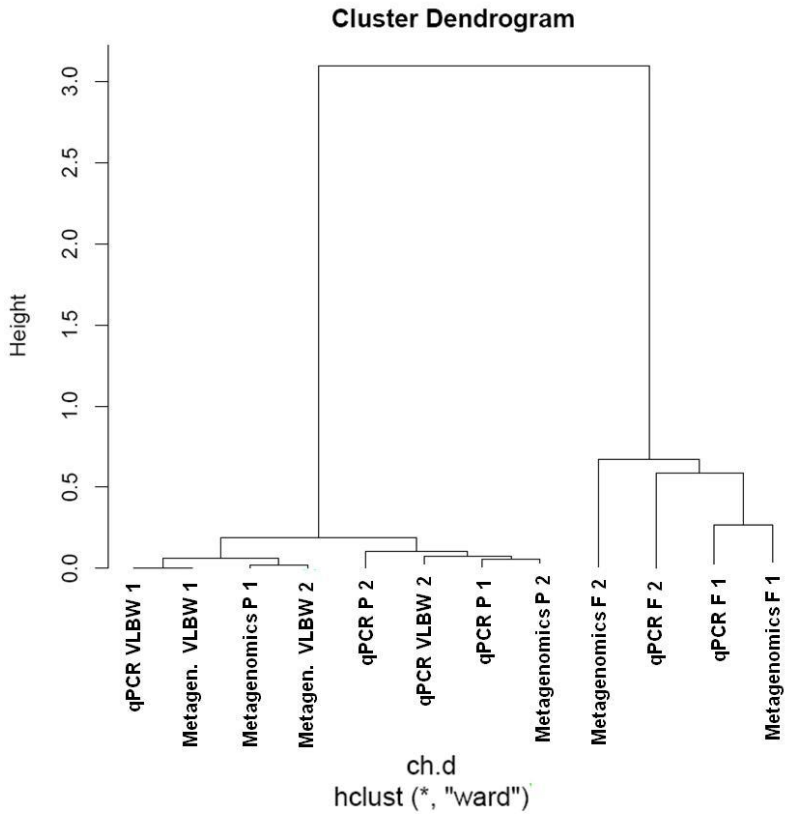

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
VLBW1
VLWB2
P1
P2
F1
F2
Other
Staphylococcaceae
Moraxellaceae
Vibrionaceae
Bacteroidaceae
Bifidobacteriaceae
Veillonellaceae
Leuconostocaceae
Enterococcaceae
Micrococcaceae
Lactobacillaceae
Clostridiaceae
Streptococcaceae
Enterobacteriaceae

Figure2

A


PCA
qpcr
metagenomics
Bifidobacterium
Enterobacteriaceae
Streptococcus
Lactobacillus
P1
P2
VLBW1
VLBW2
F1
F2
PC2(11.7%)
PC1(87.7%)
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.8
−0.6
−0.4
−0.2
0.0
0.2

B


Cluster Dendrogram
Height
3.0
2.5
2.0
1.5
1.0
0.5
0.0
qPCR VLBW 1
Metagen. VLBW 1
Metagenomics P 1
Metagen. VLBW 2
qPCR P 2
qPCR VLBW 2
qPCR P 1
Metagenomics P 2
Metagenomics F 2
qPCR F 2
qPCR F 1
Metagenomics F 1
ch.d
hclust (*, "ward")